Statistica Sinica Preprint No: SS-2019-0354	
Title	SEAMLESS PHASE II/III CLINICAL TRIALS WITH COVARIATE ADAPTIVE RANDOMIZATION
Manuscript ID	SS-2019-0354
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202019.0354
Complete List of Authors	Wei Ma, Mengxi Wang and Hongjian Zhu
Corresponding Author	Hongjian Zhu
E-mail	hongjian.zhu@uth.tmc.edu
Notice: Accepted version subject to English editing.	

SEAMLESS PHASE II/III CLINICAL TRIALS WITH COVARIATE ADAPTIVE RANDOMIZATION

Wei Ma[1], Mengxi Wang[2] and Hongjian Zhu[2*]

[1]*Renmin University of China,* [2]*The University of Texas Health Science Center at Houston*

Abstract: There is an urgent need to evaluate new therapies in a time-sensitive and cost-effective manner. We propose the adaptive seamless phase II/III clinical trials with covariate adaptive randomization (CAR) to satisfy this urgent need. CAR is one of the most popular designs in randomized controlled trials; it enhances covariance balance and ensures valid treatment comparisons. The challenges include (1) the type I error rate of the commonly used Student's t-test following CAR can be inflated because of the seamless trials but decreased with CAR; (2) the complicated allocation mechanism induced by CAR causes extra difficulties to derive the asymptotic properties of a test procedure; and (3) previous theoretical studies of seamless trials mainly rely on the assumption of complete randomization, a procedure rarely used in real trials. We establish a theoretical foundation for adaptive seamless phase II/III trials with CAR. We also propose an approach that is easy to implement to control the type I error rate and to improve the power while the Student's t-test is employed. This is an important step that will promote the application of this procedure.

Key words and phrases: Adaptive design, Type I error, Power.

1. Introduction

In 2006 the US Food and Drug Administration (FDA) emphasized the importance of streamlining clinical trials (US FDA, 2006). Since then, there has been an urgent need to evaluate new therapies in a time-sensitive and cost-effective manner without compromising the integrity and validity of the development process. In this paper, we propose the adaptive seamless phase II/III clinical trials with covariate adaptive randomization (CAR) to satisfy this urgent need. Recently, the FDA drafted guidance on seamless clinical trials, aiming to broaden acceptance of the design (US FDA, 2018). CAR is one of the most popular clinical trial designs. It ensures valid treatment comparisons by balancing the potentially confounding patient characteristics across the treatment arms. We establish a theoretical foundation for adaptive seamless phase II/III trials with CAR to facilitate the application of this design in practice. We address three major challenges: the theoretical properties of this complicated allocation and analysis procedure; control of the type I error rate; and improvement of the power.

In a typical seamless phase II/III clinical trial (Thall et al., 1988; Jennison and Turnbull, 2007; Hampson and Jennison, 2015), multiple experimental treatments or drug doses are simultaneously compared against a control in the phase II trial; the candidates with the best performance are then selected for the phase III trial; and an analysis based on data from both phases is performed at the end of the trial. With a single protocol for the two phases, the seamless design avoids the lead time between conventional phase II and phase III trials, which is likely to be six months or more. It also reduces the number of trials required to compare multiple drugs, decreases the sample size, and allows longer monitoring of the patients from phase II (Bretz et al., 2009). These advantages increase the profits of pharmaceutical companies and have received much attention from industry. By 2016, more than 40 active, first-in-human cancer trials had used the seamless strategy (Prowell et al., 2016). An example highlighted by Bhatt and Mehta (2016) is the Indacaterol to Help Achieve New COPD Treatment Excellence (INHANCE) trial Barnes et al. (2010), a seamless phase II/III clinical trial of inhaled indacaterol for the treatment of chronic obstructive pulmonary disease (COPD) using equal allocation with stratification for smoker status.

For seamless clinical trials, it is critical to control the possibly inflated type I error rate under the dual influence of multiplicity and selection (Bauer et al., 2010). Following the approach of Bauer and Kieser (1999), Bretz et al. (2006) and Schmidli et al. (2006) used the closure principle (Marcus et al., 1976), combination tests (Bauer and Köhne, 1994; Lehmacher and

Wassmer, 1999), and multiple testing procedures (Simes, 1986; Dunnett, 1955) to control the familywise type I error rate. Liu et al. (2002) provided a solid theoretical foundation for general two-stage adaptive designs. Koenig et al. (2008) proposed the adaptive Dunnett test based on the conditional error rate (Müller and Schäfer, 2001). However, the theory of most of these studies assume complete randomization with independent responses, which is rarely applied in clinical trials, and these approaches may not be valid under other randomization schemes.

It is well known that an imbalance of the confounding covariates across treatments may bias the study results, and this imbalance can be mitigated by CAR that sequentially assigns the next patient based on the previous treatment assignments and covariates as well as the current covariate profile. CAR can also reduce the selection bias, minimize the accidental bias, and improve the statistical efficiency (Shao et al., 2010). The most commonly used CAR in randomized controlled trials is the stratified permuted block (SPB) design. Other CAR designs and clinical trials adopting CAR include those of Pocock and Simon (1975), Antognini and Zagoraiou (2011), Iacono et al. (2006), Jakob et al. (2012), and Krueger et al. (2007) as well as Barnes et al. (2010) mentioned above, a seamless phase II/III trial with CAR.

In practice, unadjusted analyses, such as Student's t-test, are commonly used in clinical trials (Kahan et al., 2014; Sverdlov, 2015). This simple approach avoids model misspecification but results in a conservative type I error rate under CAR (Shao et al., 2010). Hypothesis testing and sequential monitoring in clinical trials with CAR have recently been studied by Shao et al. (2010), Ma et al. (2015), Bugni et al. (2018), and Zhu and Hu (2019). However, none of these studies investigated the application of CAR in seamless phase II/III trials.

Both seamless phase II/III designs and CAR with the Student's t-test lead to difficulties in controlling the type I error rate. It is challenging to perform the theoretical investigation and propose approaches to control the type I error rate for seamless phase II/III trials with CAR, because (1) the correlation structure of the within-stratum imbalances is complex; (2) the relationships among the treatment assignments, covariates, and responses are complicated; (3) the allocation functions are discrete; and (4) the data used in the treatment selection are also used for inference at the end of the trial. Therefore, seamless phase II/III clinical trials with CAR currently lack a theoretical foundation, and control of the type I error rate is based on the assumption of complete randomization.

In this paper, we provide a theoretical foundation for seamless phase II/III clinical trials with CAR. We also propose ways to adjust the Student's

t-test statistics and use the test procedures available for complete randomization to control the type I error rate and improve the power. Therefore, clinical trial practitioners can have valid tests and treatment comparisons in seamless clinical trials with CAR. We also investigate estimation and hypothesis testing for CAR with multiple treatments, which has a crucial implication for a single phase with multiple treatments. Our numerical studies show that compared to traditional methods our procedure can control the type I error rate well and significantly increase the power.

2. Seamless phase II/III clinical trials with CAR

2.1 Framework of seamless phase II/III clinical trials

We consider a seamless phase II/III trial and refer to the two phases as Stage 1 and Stage 2. Assume the planned sample size for Stage 1 is N, and the planned sample size for Stage 2 is N', so the total sample size is $N + N'$. The *design procedure* is described below.

Stage 1. The first N patients are sequentially assigned to K experimental treatments and the control arm with CAR. One treatment, say treatment $k^\star$, is then chosen for Stage 2 based on certain criteria, e.g., the one with the largest estimated treatment effect and an acceptable safety profile.

Stage 2. The remaining N' patients are sequentially assigned to treatment $k^\star$ and the control arm with CAR. A final analysis comparing treatment $k^\star$ and the control arm is performed using the data from both stages.

We next describe the *analysis procedure* with a flowchart in Figure 1.

Let $\boldsymbol{\mu} = (\mu_0, \mu_1, \ldots, \mu_K)^{\mathrm{T}}$ denote the vector of treatment effects, with μ_0 corresponding to the control arm and $\mu_k, k = 1, \ldots, K$, corresponding to K experimental treatments. At the end of the trial, without loss of generality, we test $H_{0,k^\star} : \mu_{k^\star} = \mu_0$ vs. $H_{1,k^\star} : \mu_{k^\star} > \mu_0$ based on the combined data from the two stages with the closure principle (Marcus et al., 1976) to control the familywise type I error rate. The closure principle rejects $H_{0,k^\star}$ at level α if each intersection hypothesis $H_{0,I}$ with $k^\star \in I, I \subseteq \{1, \ldots, K\}$, is rejected at level α, where $H_{0,I} = \cap_{k \in I} H_{0,k}$ with $H_{0,k} : \mu_k = \mu_0$.

To test each intersection hypothesis $H_{0,I}$ using the data from two stages, we use a combination test such as the inverse χ^2 method (Bauer and Köhne, 1994). Let $P_{1,I}$ and $P_{2,I}$ denote the p-values for $H_{0,I}$ based on the data from Stage 1 and Stage 2, respectively. Then the inverse χ^2 method rejects $H_{0,I}$ if $-\log(P_{1,I}P_{2,I}) > \chi^2_4(1-\alpha)/2$, where $\chi^2_4(1-\alpha)$ is the $(1-\alpha)$th quantile of the χ^2 distribution with 4 degrees of freedom. An alternative approach is the weighted inverse normal method (Lehmacher and Wassmer, 1999).

To perform the combination test, we calculate the adjusted p-values for each stage, $P_{1,I}$ and $P_{2,I}$, using either the Simes test or the Dunnett test. Note that both these two tests reduce to the usual Student's t-test if there are only one treatment and one control arm, as in Stage 2. We now briefly review the Simes test and the Dunnett test when they are used under complete randomization, deferring the justification and modification of these methods under CAR to Section 2.3. We illustrate the test procedures for Stage 1 with multiple treatments since the Stage 2 comparison between two arms is straightforward.

Suppose the intersection hypothesis $H_{0,I}$ is composed of m elementary hypotheses $H_{0,k}$ with the associated p-values denoted by $P_{1,k}$. Let $P_{1,(j)}, j = 1, \ldots, m$, be the p-values in ascending order. Using the Simes test, we have the adjusted p-value $P_{1,I} = \min_{1\le j\le m}\big(mP_{1,(j)}/j\big)$ for the intersection hypothesis $H_{0,I}$.

For the Dunnett test, without loss of generality, consider $H_{0,I}$ with $I = \{1, \ldots, K\}$. Let

$$t_k = (\bar{Y}_k - \bar{Y}_0)/\big\{s(1/N_k + 1/N_0)^{1/2}\big\}, \quad k = 1, \ldots, K, \tag{2.1}$$

where N_k is the number of patients assigned to treatment k; $\bar{Y}_k$ and S_k^2 are the sample mean and sample variance, respectively, under treatment k; and $s^2 = \sum_{k=0}^{K}(N_k-1)S_k^2/\nu$ with $\nu = N-K-1$. Under complete randomization the null distribution of $(t_1, \ldots, t_K)^{\mathrm{T}}$ is the K-variate t-distribution with ν

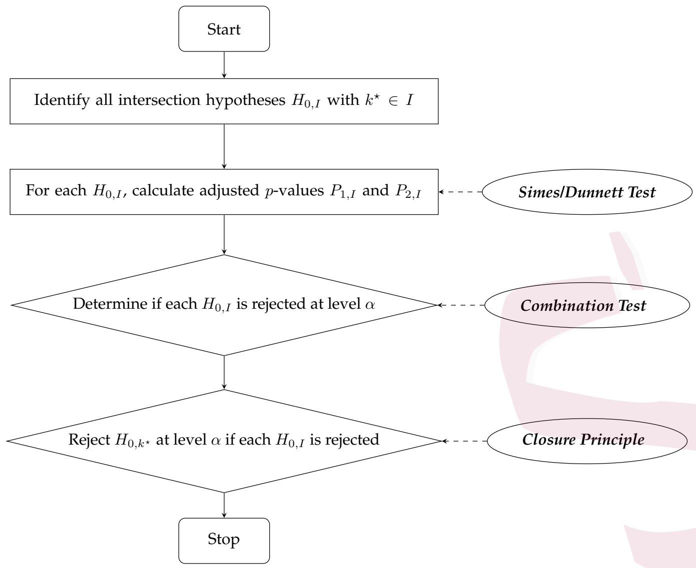

Figure 1: Flowchart of the analysis procedure of testing $H_{0,k^\star}$.

degrees of freedom and correlations

$$\rho_{k,k'} = \left(\frac{N_k}{N_k + N_0}\right)^{1/2} \left(\frac{N_{k'}}{N_{k'} + N_0}\right)^{1/2}, \quad k, k' = 1, \ldots, K.$$

Then the conventional Dunnett test rejects the intersection hypothesis $H_{0,I}$ at level α if

$$\max_{1 \le k \le K} t_k \ge c,$$

where c is determined by $\mathrm{pr}(\zeta_1 < c, , \ldots, \zeta_K < c) = 1 - \alpha$, and $(\zeta_1, \ldots, \zeta_K)^{\mathrm{T}}$ follows the K-variate t-distribution with ν degrees of freedom and correlations $\rho_{k,k'}$.

In the literature, the above analysis procedure is used in seamless trials to control the familywise type I error rate with the assumption that the patients are allocated by complete randomization and the responses of the patients are independent of each other. However, the responses and treatment assignments are no longer independent under CAR, because of the complicated randomization mechanism that balances covariates over different arms. When there are two arms in a phase III clinical trial, the conventional tests are too conservative with a small type I error rate because of CAR (Shao et al., 2010; Ma et al., 2015). It is unclear whether CAR will lead to a conservative type I error rate in seamless clinical trials, and it is worth investigating the underlying theory. Based on the closure principle (Marcus et al., 1976) and the conditional invariance principle (Brannath et al., 2007, 2012), for a valid treatment comparison it suffices to validate the Simes test and the Dunnett test under CAR for each stage of the above design and analysis procedure.

2.2 Estimation following CAR with multiple treatments

In this section, we study estimation for CAR with multiple treatments, a key element for an adaptive seamless II/III trial and an important problem in its own right with implications for a traditional single-phase clinical trial with CAR and multiple treatments.

Suppose a CAR procedure is implemented to assign the patients to $(K+1)$ arms, and the total sample size is N. Let $Z_i, i = 1, \ldots, N$, represent the covariate information for the ith patient. We allow Z_i to be either discrete or continuous covariates and assume that all the covariates are independent and identically distributed. To incorporate continuous covariates into the randomization procedure, we discretize Z_i with $D(Z_i)$, a discrete function of Z_i taking values in a finite set $\mathcal{D}$. We can set $D(Z_i) = Z_i$ for discrete covariates, so both types of covariates can be treated using the same notation. For simplicity, we introduce our methods using the univariate covariate Z_i with variance σ_z^2; the conclusions can be easily extended to multivariate cases.

Let $\boldsymbol{T_i} = (T_{i0}, T_{i1}, \ldots, T_{iK})^{\mathrm{T}}$ indicate the treatment assignment for the ith patient, where treatment 0 represents the control arm. We have $T_{ik} = 1, k = 0, 1, \ldots, K$, if the ith patient is assigned to treatment k, and $T_{ik} = 0$ otherwise. Then $N_k = \sum_{i=1}^{N} T_{ik}, k = 0, 1, \ldots, K$, is the number of patients in treatment k after N patients have been assigned. Let $\boldsymbol{Y_i} = (Y_{i0}, Y_{i1}, \ldots, Y_{iK})^{\mathrm{T}}, i = 1, \ldots, N$, be a random vector of response variables, where $Y_{ik}, k = 0, 1, \ldots, K$, is the response of the ith patient under treatment k. Only one element of $\boldsymbol{Y_i}$, say Y_{ik}, can be observed if $T_{ik} = 1$.

Assume the response of the ith patient under treatment k follows

$$Y_{ik} = \mu_k + \beta Z_i + \varepsilon_{ik}, \quad i = 1, \ldots, N,$$

where β represents the covariate effect, and the ε_{ik}s are independent and identically distributed random errors with mean zero and constant variance σ_ε^2 and are independent of the covariates. In practice, to avoid unnecessary or incorrect model assumptions, a natural treatment effect estimator for treatment $k, k = 0, 1, \ldots, K$, is $\bar{Y}_k = \sum_{i=1}^{N} T_{ik} Y_{ik} / N_k$.

We first introduce two conditions for the balancing properties under CAR with multiple treatments. For any $k = 1, \ldots, K$,

Condition A. $N_k - N_0 = O_p(1)$.

Condition B. $\sum_{i=1}^{N} (T_{ik} - T_{i0}) I\{D(Z_i) = d\} = O_p(1)$ for any $d \in \mathcal{D}$.

These two conditions ensure that good balancing properties are attained under a CAR procedure. Condition (A) indicates that the number of patients in each treatment group is approximately equal, and Condition (B) implies a balance of treatment assignments within each covariate stratum formed by $D(Z_i)$. Both conditions are satisfied by the stratified permuted block design with multiple treatments. Note that Condition (B) implies Condition (A) when the number of stratum is finite. We list both conditions to emphasize the balancing properties with respect to different levels (overall and within-stratum), similarly as in Ma et al. (2015).

Remark 1. The two conditions can be considered a generalization of those used in Shao et al. (2010) and Ma et al. (2015), where only two arms (one treatment and one control) are considered.

Now we present our theorem regarding the treatment effect estimation. We write $\mathbf{1}$ for a column vector of ones, with the subscript denoting its dimension.

Theorem 1. *Under Conditions (A) and (B), as $N \to \infty$,*

$$\left(\frac{N}{K+1}\right)^{1/2}\left\{\left(\bar{Y}_0, \bar{Y}_1, \ldots, \bar{Y}_K\right)^{\mathrm{T}} - \left(\mu_0, \mu_1, \ldots, \mu_K\right)^{\mathrm{T}}\right\}$$

converges in distribution to a normal distribution with mean zero and covariance matrix $\mathbf{V}$, where $\mathbf{V} = \mathrm{diag}\{\sigma_d^2 \mathbf{1}_{K+1}\} + (K+1)^{-1}\beta^2 \mathrm{Var}[E\{Z_i \mid D(Z_i)\}]\mathbf{1}_{K+1}\mathbf{1}_{K+1}^{\mathrm{T}}$, and $\sigma_d^2 = \sigma_\varepsilon^2 + \beta^2 E[\mathrm{Var}\{Z_i \mid D(Z_i)\}]$.

The theorem gives the asymptotic distribution of the average responses of different treatment groups. It is clear that these treatment effect estimators are no longer independent and are positively correlated, which is a key difference compared to complete randomization. The dependence structure arises from the randomization procedure that adaptively assigns patients to the treatment arms to enhance the covariate balance.

Remark 2. Under complete randomization, the asymptotic covariance matrix of $\{N/(K+1)\}^{1/2}(\bar{Y}_0, \bar{Y}_1, \ldots, \bar{Y}_K)^{\mathrm{T}}$ is a diagonal matrix with the diagonal entries equal to $\sigma_\varepsilon^2+\beta^2\sigma_Z^2$, which is larger than $\sigma_d^2+(K+1)^{-1}\beta^2\mathrm{Var}[E\{Z_i \mid D(Z_i)\}]$ under CAR. Thus, CAR can increase the precision of the estimation of the mean response of each treatment group by balancing the covariates.

The theorem can be used to study the properties of any linear transformation of $(\bar{Y}_0, \bar{Y}_1, \ldots, \bar{Y}_K)^{\mathrm{T}}$. However, our main interest is in comparing the treatment effects between the experimental treatments and the control. The next corollary is a direct consequence of Theorem 1 and provides the asymptotic joint distribution of $\bar{Y}_k - \bar{Y}_0$, $k = 1, \ldots, K$.

Corollary 1. *Under Conditions (A) and (B), as $N \to \infty$,*

$$\left(\frac{N}{K+1}\right)^{1/2}\left\{\left(\bar{Y}_1 - \bar{Y}_0, \ldots, \bar{Y}_K - \bar{Y}_0\right)^{\mathrm{T}} - (\mu_1 - \mu_0, \ldots, \mu_K - \mu_0)^{\mathrm{T}}\right\}$$

converges in distribution to a normal distribution with mean zero and covariance matrix $\boldsymbol{\Sigma}$, where $\boldsymbol{\Sigma} = \mathrm{diag}\{\sigma_d^2\mathbf{1}_K\} + \sigma_d^2\mathbf{1}_K\mathbf{1}_K^{\mathrm{T}}$.

Corollary 1 reveals that the asymptotic variance of $\bar{Y}_k - \bar{Y}_0$ under CAR is smaller than that under complete randomization. In particular, when the Z_is are discrete covariates, the asymptotic variance of $\{N/(K+1)\}^{1/2}(\bar{Y}_k - \bar{Y}_0)$ is $2\sigma_\varepsilon^2$, compared to $2(\sigma_\varepsilon^2 + \beta^2\sigma_z^2)$ under complete randomization. This can be interpreted to mean that the covariates are balanced so well that the variability of the difference in means between the two groups is due only to the random errors. The corollary provides a theoretical foundation for deriving a valid test with a correct type I error rate.

2.3 Control of type I error rate in seamless clinical trials with CAR

The commonly used test statistic for $H_{0,k} : \mu_k = \mu_0$ is based on a form of $\bar{Y}_k - \bar{Y}_0$ that is normalized to have a unit variance. The next theorem follows Theorem 1 and shows how to construct such test statistics.

Theorem 2. *Assume that Conditions (A) and (B) hold. Let*

$$X_k = (\bar{Y}_k - \bar{Y}_0)/\{\sigma_d(1/N_k + 1/N_0)^{1/2}\}, \quad k = 1, \ldots, K.$$

If the null hypotheses $H_{0,k} : \mu_k = \mu_0$ are true for all $k = 1, \ldots, K$, then, as $N \to \infty$, $(X_1, \ldots, X_K)^{\mathrm{T}}$ converges in distribution to a normal distribution with mean zero and covariance matrix $\mathbf{R}$, where $\mathbf{R} = \mathrm{diag}\{\mathbf{1}_K/2\} + \mathbf{1}_K\mathbf{1}_K^{\mathrm{T}}/2$.

Based on Theorem 2, X_k following a standard normal distribution can be used as the test statistic to test the individual null hypothesis $H_{0,k} : \mu_k = \mu_0$, and the critical value can be selected accordingly. Note that the asymptotic distribution remains unchanged if σ_d is replaced by its consistent estimator $\hat{\sigma}_d$, which is usually obtained in practice by either the model-based method or the bootstrap method. We propose to fit a linear regression with all the stratification covariates in the model to obtain consistent estimators for the parameters in the expression of σ_d in Theorem 1 and to calculate the estimate of σ_d accordingly. By the continuous mapping theorem, $\hat{\sigma}_d$ obtained in this way is a consistent estimator of σ_d. We illustrate these methods in Section 3.

Remark 3. Compared to t_k defined in (2.1) that is valid under complete randomization, we find that σ_d or its consistent estimator must be used instead of s to construct the test statistics under CAR. Otherwise, the asymptotic distribution is more concentrated around zero than the standard normal distribution, and the actual type I error rates are smaller than the nominal levels.

As argued before, to control the type I error rate for seamless phase II/III clinical trials, it is critical and sufficient to prove that the Simes test or the Dunnett test is still valid with the test statistics X_k under CAR. In Theorem 2 we have successfully detected that the joint distribution of $(X_1, ..., X_K)^{\mathrm{T}}$ is an equicorrelated multivariate normal with non-negative correlation. The following theorem is an immediate consequence of Result 1 in Sarkar and Chang (1997).

Theorem 3. *Under Conditions (A) and (B), the type I error rate is controlled for the Simes test with the test statistics $X_k, k = 1, \ldots, K$, under CAR.*

We next consider the Dunnett test. In Theorem 2, we proved that the vector of test statistics $(X_1, \ldots, X_K)^{\mathrm{T}}$ asymptotically follows a K-dimensional normal distribution with unit variances and constant correlations equal to 1/2. To obtain a valid test, we can reject the null hypotheses if

$$\max_{1 \leq k \leq K} X_k \geq c', \tag{2.2}$$

where c' is determined by $\mathrm{pr}(\xi_1 < c', \ldots, \xi_K < c') = 1 - \alpha$ and $(\xi_1, \ldots, \xi_K)^{\mathrm{T}}$ follows the normal distribution $\mathcal{N}(\mathbf{0}, \mathbf{R})$. It is important to note that the test considered here is based on X_k defined in Theorem 2, instead of the conventional t_k that is used under complete randomization. Also, the original Dunnett test is based on the multivariate t distribution, while the test presented here uses the normal distribution, which relies on the asymptotic normality given in Theorem 2. For these reasons, we refer to the test based on X_k and rejection region (2.2) as the modified Dunnett test, although we call it the Dunnett test for simplicity when there is no confusion.

An application of Theorem 2 yields the following theorem.

Theorem 4. *Under Conditions (A) and (B), the type I error rate is asymptotically α for the Dunnett test with test statistics $X_k, k = 1, \ldots, K$, under CAR.*

Theorems 3 and 4 show that the widely used Simes and Dunnett tests can also be applied under CAR, provided an appropriate adjustment is made to the test statistics. Combined with the results from the last section, the design and analysis procedures for seamless phase II/III clinical trials with CAR (described in Section 2.1) can lead to higher precision and valid inference for treatment effects, showing the advantages of balancing covariates over complete randomization.

3. Numerical studies

We have obtained the asymptotic results for the proposed procedure. We next study its finite-sample properties regarding the type I error rate, the power, and the probability that the best treatment is selected for Stage 2 at the interim look. Three scenarios are considered: (1) three treatments and two stratification covariates; (2) four treatments and three stratification covariates; (3) five treatments and two stratification covariates. We study both discrete and continuous stratification covariates. In this section, we discuss the simulation setting and results of Scenario 1. Results of Scenarios 2 and 3 and additional results showing the robustness of the proposed method to various model misspecification are also reported in the supplementary materials.

We first consider the case of discrete stratification covariates. In Scenario 1, two experimental treatments (i.e., treatment 1 and treatment 2) are compared with one control (i.e., treatment 0) in Stage 1, and discrete stratification covariates are considered. The following linear model with two covariates Z_1 and Z_2 is used to simulate response $Y_i, i = 1, \ldots, N + N'$,

$$Y_i = \alpha_0 + \alpha_1 T_{i1} + \alpha_2 T_{i2} + \beta_1 Z_{i1} + \beta_2 Z_{i2} + \varepsilon_i,$$

where $(\alpha_0, \alpha_1, \alpha_2, \beta_1, \beta_2)^{\mathrm{T}}$ are unknown parameters; Z_1 and Z_2 follow Bernoulli distributions with success rates p_1 and p_2, respectively; ε_i follows the normal distribution $\mathcal{N}(0, \sigma^2)$; and $T_{ik} = 1, k = 1, 2$, if the ith patient is assigned to experimental treatment k, and $T_{ik} = 0$ otherwise.

In Stage 1, 120 patients sequentially enter the trial. We implemented and compared the stratified permuted block design with respect to both Z_1 and Z_2 with a block size of 6 and complete randomization. Let

$$W_k = (\bar{Y}_k - \bar{Y}_0)/(S_k^2/N_k + S_0^2/N_0)^{1/2}, \quad k = 1, \ldots, K.$$

The experimental treatment with a larger W_k, denoted treatment $k^\star$, is considered more effective and selected to continue to Stage 2. In Stage 2, 500 patients sequentially enter the trial and are randomly allocated to the control arm and treatment $k^\star$ with either a stratified permuted block design or complete randomization. At the end of the trial, we test $H_{0,k^\star} : \mu_{k^\star} = \mu_0$ vs. $H_{1,k^\star} : \mu_{k^\star} > \mu_0$.

We compare four analysis approaches: the traditional two-sample t-test without adjustment; linear regression with both covariates Z_1 and Z_2 in the model; the bootstrap t-test proposed by Shao et al. (2010); and our t-test with adjustment. Here, we show the bootstrap t-test for Stage 1, and it can be done similarly for Stage 2. B bootstrap samples $(Y_1^{\star b}, Z_{1,1}^{\star b}, Z_{1,2}^{\star b}), \ldots, (Y_N^{\star b}, Z_{N,1}^{\star b}, Z_{N,2}^{\star b}), b = 1, 2, \ldots, B$, are generated independently by random sampling with replacement from $(Y_1, Z_{1,1}, Z_{1,2}), \ldots, (Y_N, Z_{N,1}, Z_{N,2})$. We implement stratified permuted block design randomization with respect to $(Z_{1,1}^{\star b}, Z_{1,2}^{\star b}), \ldots, (Z_{N,1}^{\star b}, Z_{N,2}^{\star b})$ to obtain the bootstrap analogs of treatment allocations $(T_{1k}^{\star b}, \ldots, T_{Nk}^{\star b})$, where $T_{ik}^{\star b} = 1, k = 0, 1, 2$, if the ith patient is assigned to treatment k, and $T_{ik}^{\star b} = 0$ otherwise. Define

$$\bar{Y}_k^{\star b} - \bar{Y}_0^{\star b} = \frac{1}{N_k^{\star b}} \sum_{i=1}^{N} T_{ik}^{\star b} Y_i^{\star b} - \frac{1}{N_0^{\star}} \sum_{i=1}^{N} T_{i0}^{\star b} Y_i^{\star b},$$

$$N_0^{\star b} = \sum_{i=1}^{N} T_{i0}^{\star b}, \quad N_k^{\star b} = \sum_{i=1}^{N} T_{ik}^{\star b}, \quad k = 1, 2.$$

The bootstrap estimator of the variance of $\bar{Y}_k - \bar{Y}_0$ is the sample variance of $\bar{Y}_k^{\star b} - \bar{Y}_0^{\star b}, b = 1, 2, \ldots, B$, denoted $\hat{\nu}_{Bj}$. The bootstrap t-test has the test statistic $T_B = (\bar{Y}_k - \bar{Y}_0)/\hat{\nu}_{Bj}^{1/2}$. We set $B = 200$ in the simulations. For the proposed t-test with adjustment, based on our theorems, the value of σ_d is estimated by Theorem 1, and the values of σ_ε and β are obtained by fitting a linear model with both covariates. The closure principle and a combination test with either the Simes or Dunnett test are applied to control the familywise type I error rate. The significance level α is 0.05 for all the tests. All the results are based on $10,000$ replications.

In Table 1, we report the type I error rate for different parameter values of (p_1, p_2, σ) while fixing $\alpha_0 = \beta_1 = \beta_2 = 1$. Under complete randomization, the type I error rate is close to the nominal level 0.05 for both the two-sample t-test (*t-test*) and the full linear model (*lm*). Under the SPB design with either the Dunnett or Simes test, the type I error rate of the two-sample t-test is far below 0.05, while our t-test with adjustment (*Adjusted-t*) successfully controls the error rate. The error is also well controlled when we use the full linear model or the bootstrap t-test (*BS-t*).

In Table 2 we compare the power of different designs and analysis approaches. We report results for different values of (α_1, α_2) while fixing $(p_1, p_2, \sigma) = (0.5, 0.5, 1)$ and $\alpha_0 = \beta_1 = \beta_2 = 1$. Our t-test with adjustment and the bootstrap t-test under CAR have a significantly higher power than

Table 1: Type I error rate (percentage) in seamless trial with three treatments and two discrete covariates.

	(p_1, p_2, σ)	Allocation	*t-test*	*lm*	*BS-t*	*Adjusted-t*
Simes	$(0.5, 0.5, 1.0)$	SPB	1.73	5.26	5.14	5.20
		CR	5.00	4.73	-	-
	$(0.4, 0.6, 1.0)$	SPB	1.78	4.84	5.35	5.41
		CR	4.73	4.80	-	-
	$(0.4, 0.6, 1.5)$	SPB	3.00	4.78	5.46	5.36
		CR	4.61	4.65	-	-
Dunnett	$(0.5, 0.5, 1.0)$	SPB	1.98	5.75	5.09	5.46
		CR	5.20	5.30	-	-
	$(0.4, 0.6, 1.0)$	SPB	1.91	5.38	5.23	5.36
		CR	5.05	5.23	-	-
	$(0.4, 0.6, 1.5)$	SPB	3.38	5.27	5.17	5.40
		CR	5.09	5.08	-	-

the t-test without adjustment under either CAR or complete randomization. In addition, our design performs better than complete randomization in terms of the number of replications (M) in which the better treatment is selected for Stage 2. To save space, we leave additional results for Tables 2 and 4 in the Supplement.

Table 2: Power (percentage) and number (M) of replications in which the better treatment is selected for Stage 2 in seamless trial with three treatments and two discrete covariates.

	(α_1, α_2)	Allocation	*t-test*	*lm*	*BS-t*	*Adjusted-t*	M
Simes	$(0.26, 0.16)$	SPB	65.11	79.88	80.48	80.55	6667
		CR	64.83	79.25	-	-	6420
	$(0.24, 0.16)$	SPB	58.96	75.35	76.49	76.42	6374
		CR	60.27	74.76	-	-	6139
	$(0.22, 0.16)$	SPB	52.69	70.23	71.61	71.33	6042
		CR	55.07	69.79	-	-	5837
Dunnett	$(0.26, 0.16)$	SPB	65.74	80.61	80.86	80.97	6667
		CR	65.98	80.13	-	-	6420
	$(0.24, 0.16)$	SPB	60.08	76.30	77.20	77.00	6374
		CR	61.44	75.82	-	-	6139
	$(0.22, 0.16)$	SPB	53.57	71.18	72.39	72.10	6042
		CR	56.28	71.09	-	-	5837

We also performed numerical studies in which some of the covariates are continuous. To save space, we report results only for three treatments and two stratification covariates. The setting is as in Scenario 1 except that we assume Z_2 follows the standard normal distribution. When implementing the stratified permuted block design, we discretize Z_2 into the Bernoulli variable $D(Z_2)$ as follows, $D(Z_2) = 1$ if $Z_2 < z_q$, and $D(Z_2) = 0$ otherwise, where z_q is the qth quantile of the standard normal distribution. The continuous covariate is used in the statistical inference procedures. Our t-test with adjustment controls the type I error at around 0.05, while the two-sample t-test is too conservative under the SPB design design with either the Dunnett or Simes test (Table 3). Meanwhile, the t-test with adjustment is much more powerful than the two-sample t-test under both the stratified permuted block design and complete randomization (Table 4).

4. Redesign of a clinical trial evaluating treatments for chronic obstructive pulmonary disease

COPD is a chronic lung inflammation disease that causes poor airflow from the lungs and long-term breathing problems. A double-blinded two-stage seamless clinical trial, also known as the Indacaterol to Help Achieve New COPD Treatment Excellence (INHANCE) trial, has been conducted to evaluate the efficacy and safety of indacaterol in the treatment of COPD. The trial used equal allocation with stratification for smoking status (Barnes et al., 2010; Donohue et al., 2010). In Stage 1, 770 patients were enrolled and four doses of indacaterol were compared to a placebo and to two active

Table 3: Type I error rate (percentage) in seamless trial with three treatments, one discrete covariate, and one continuous covariate.

	(p_1, q, σ^2)	Allocation	*t-test*	*lm*	*BS-t*	*Adjusted-t*
Simes	$(0.5, 0.5, 1.0)$	SPB	1.10	4.53	5.45	5.16
		CR	4.47	4.56	-	-
	$(0.4, 0.6, 1.0)$	SPB	1.08	4.63	5.14	5.20
		CR	4.57	4.60	-	-
	$(0.4, 0.6, 1.5)$	SPB	2.16	4.89	5.31	4.96
		CR	4.55	4.58	-	-
Dunnett	$(0.5, 0.5, 1.0)$	SPB	1.23	4.89	5.78	5.41
		CR	5.02	4.97	-	-
	$(0.4, 0.6, 1.0)$	SPB	1.27	4.94	5.46	5.19
		CR	5.13	4.87	-	-
	$(0.4, 0.6, 1.5)$	SPB	2.31	5.09	5.66	5.31
		CR	4.89	5.10	-	-

controls, formoterol and tiotropium. In Stage 2, two doses of indacaterol were selected for comparison with placebo and tiotropium in 1683 patients.

Here we redesign the INHANCE trial and evaluate the differences of trough forced expiratory volume in one second (FEV1) between multiple doses of indacaterol and the placebo. Trough FEV1 is a standard measure-

Table 4: Power (percentage) and number (M) of replications in which the better treatment is selected for Stage 2 in seamless trial with three treatments, one discrete covariate, and one continuous covariate.

	(α_1, α_2)	Allocation	*t-test*	*lm*	*BS-t*	*Adjusted-t*	M
Simes	$(0.26, 0.16)$	SPB	46.76	79.90	69.52	69.61	6547
		CR	49.92	79.16	-	-	6154
	$(0.24, 0.16)$	SPB	40.55	75.70	64.68	64.97	6243
		CR	45.60	74.73	-	-	5970
	$(0.22, 0.16)$	SPB	35.35	70.98	60.07	59.83	5944
		CR	41.69	70.32	-	-	5709
Dunnett	$(0.26, 0.16)$	SPB	48.14	80.49	70.19	70.18	6547
		CR	51.38	79.77	-	-	6154
	$(0.24, 0.16)$	SPB	41.79	76.43	65.67	65.87	6243
		CR	46.87	75.79	-	-	5970
	$(0.22, 0.16)$	SPB	36.79	72.06	60.77	60.68	5944
		CR	43.15	71.29	-	-	5709

ment of lung capacity, and lower FEV1 indicates more severe COPD. We simplify the treatment arms into the placebo and four dose levels of indacaterol in Stage 1, and we select only one dose level along with the placebo to go forward to Stage 2, following the selection rule described in the previous section. We use summary statistics for the patients and the effect sizes of the dosages in the study to create a synthetic dataset of patients by simulating the outcome FEV1 and the covariate in 112 patients in Stage 1 and 167 patients in Stage 2 through the following linear regression model:

$$Y_i = 0.15 + 0.15T_{i1} + 0.18T_{i2} + 0.22T_{i3} + 0.19T_{i4} + \beta_1 Z_{i1} + \varepsilon_i.$$

Here $(T_{i1}, T_{i2}, T_{i3}, T_{i4})$ are indicator variables indicating the dosage assignment of the ith patient: $T_{ik} = 1$, $k = 1, ..., 4$, if the ith patient is assigned to dosage k, and $T_{ik} = 0$ otherwise. The binary covariate Z_{i1} indicates smoking status with a success rate of 0.41: $Z_{i1} = 1$ if the ith patient is a current smoker and $Z_{i1} = 0$ if the patient is an ex-smoker. ε_i follows the normal distribution $\mathcal{N}(0, \sigma^2)$.

In both stages, a stratified permuted block design with respect to smoking status is implemented with block sizes of 10 and 6, respectively, to assign patients to different arms. In Table 5 we compare the power of the two-sample t-test and our t-test with adjustment using different values of β_1 and σ. We find that, with an increase in the value of the smoking status coefficient, the power advantage of our t-test with adjustment becomes more pronounced, indicating that our t-test with adjustment is especially useful when the outcome has large differences among strata that are generated by dividing the study population with stratification covariates. We also find

Table 5: Power (percentage) in redesigned INHANCE trial

	(β_1, σ)	*t-test*	*Adjusted-t*		(β_1, σ)	*t-test*	*Adjusted-t*
Simes	$(0.2, 0.5)$	83.33	85.68	Dunnett	$(0.2, 0.5)$	85.20	86.92
	$(0.6, 0.5)$	73.22	85.72		$(0.6, 0.5)$	76.13	86.87
	$(1.0, 0.5)$	53.11	85.76		$(1.0, 0.5)$	56.21	86.86
	$(0.2, 0.6)$	69.57	72.66		$(0.2, 0.6)$	72.34	74.64
	$(0.3, 0.6)$	68.34	72.59		$(0.3, 0.6)$	70.97	74.49
	$(0.2, 0.7)$	58.07	61.22		$(0.2, 0.7)$	61.66	63.15
	$(0.3, 0.7)$	56.80	61.22		$(0.3, 0.7)$	60.59	63.19

that a larger σ leads to a lower power for all the tests.

5. Conclusion

Several future research directions are of interest. First, we assumed the linear model for data generation and equal allocation probabilities to investigate the treatment effect estimators based on the differences in sample means under CAR. Some recent studies, however, indicate that the linearity and equal allocation assumptions may be relaxed. When there are only two arms (one treatment and one control), Ma et al. (2020) showed that the difference-in-means estimator is unbiased and as efficient as regression-based estimators under stratified permuted block design, even if the linear model is arbitrarily misspecified and allocation probabilities are unequal for different arms. For the case of multiple treatments, which is more relevant to seamless trials, the theoretical properties of difference-in-means estimators have not been established, though some regression-based estimators have been studied (Bugni et al., 2019). Moreover, robust variance estimators are required for valid tests under these relaxed assumptions. The usual ordinary least squares variance estimator and Huber-White sandwich estimator are valid in a two-arm trial with equal allocation, but in general, especially for unequal allocation, model-based variance estimator typically tend to fail and consistent nonparametric estimators are preferred (Bugni et al., 2018; Ma et al., 2020).

Second, estimation is often an important but secondary target for seamless phase II/III trials (Posch et al., 2005; Bowden and Glimm, 2014). We have focused on hypothesis testing, the primary concern in seamless trials and another element of statistical inference. It would be interesting to explore the bias in the estimation following our design.

Third, Stallard and Friede (2008) investigated scenarios where more than one experimental treatment continues beyond the interim analysis, and sequential monitoring is implemented in Stage 2. Magirr et al. (2012) proposed methods for any number of treatment arms, stages, and patients per treatment per stage in such trials. Investigation of these scenarios, especially group sequential monitoring at phase III, will be of particular interest to practitioners.

Fourth, work on seamless phase II/III designs and adaptive randomization under the Bayesian framework includes but is not limited to Huang et al. (2009), Yuan et al. (2011), Inoue et al. (2002), Berry (2012), and Zang and Lee (2014). These designs provided insight into our study.

Fifth, seamless phase II/III designs with different study endpoints in the two stages have been investigated (Huang et al., 2009), which has profound implications for real trials with a primary endpoint that is observed only after long-term follow-up. It is necessary to select the treatment at the interim look based on correlated short-term endpoint data. Implementing our design in this scenario will broaden its application in practice. We leave these topics as future work.

Supplementary Materials

The supplementary materials contain the proof of the main theorem, and additional simulation results.

Acknowledgments

We are grateful to the associate editor and the referees for their many helpful comments. This work is partly supported by grant 11801559 from the National Natural Science Foundation of China and grant DMS-2014951 from the National Science Foundation (USA).

References

Antognini, A. B. and M. Zagoraiou (2011). The covariate-adaptive biased coin design for balancing clinical trials in the presence of prognostic factors. *Biometrika 98*(3), 519–535.

Barnes, P. J., S. J. Pocock, H. Magnussen, A. Iqbal, B. Kramer, M. Higgins, and D. Lawrence (2010). Integrating indacaterol dose selection in a clinical study in COPD using an adaptive seamless design. *Pulmonary Pharmacology and Therapeutics 23*(3), 165–171.

Bauer, P. and M. Kieser (1999). Combining different phases in the development of medical treatments within a single trial. *Statistics in Medicine 18*(14), 1833–1848.

Bauer, P., F. Koenig, W. Brannath, and M. Posch (2010). Selection and bias–Two hostile brothers. *Statistics in Medicine 29*(1), 1–13.

Bauer, P. and K. Köhne (1994). Evaluation of experiments with adaptive interim analyses. *Biometrics 50*(4), 1029–1041.

Berry, D. A. (2012). Adaptive clinical trials in oncology. *Nature Reviews Clinical Oncology 9*(4), 199–207.

Bhatt, D. L. and C. Mehta (2016). Adaptive designs for clinical trials. *New England Journal of Medicine 375*(1), 65–74.

Bowden, J. and E. Glimm (2014). Conditionally unbiased and near unbiased estimation of the selected treatment mean for multistage drop-the-losers trials. *Biometrical Journal 56*(2), 332–349.

Brannath, W., G. Gutjahr, and P. Bauer (2012). Probabilistic foundation of confirmatory adaptive designs. *Journal of the American Statistical Association 107*(498), 824–832.

Brannath, W., F. Koenig, and P. Bauer (2007). Multiplicity and flexibility in clinical trials. *Pharmaceutical Statistics: The Journal of Applied Statistics in the Pharmaceutical Industry 6*(3), 205–216.

Bretz, F., F. Koenig, W. Brannath, E. Glimm, and M. Posch (2009). Adaptive designs for confirmatory clinical trials. *Statistics in Medicine 28*(8), 1181–1217.

Bretz, F., H. Schmidli, F. König, A. Racine, and W. Maurer (2006). Confirmatory seamless phase II/III clinical trials with hypotheses selection at interim: General concepts. *Biometrical Journal 48*(4), 623–634.

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

Bugni, F. A., I. A. Canay, and A. M. Shaikh (2019). Inference under covariate-adaptive randomization with multiple treatments. *Quantitative Economics 10*(4), 1747–1785.

Donohue, J. F., C. Fogarty, J. Lotvall, D. A. Mahler, H. Worth, A. Yorgancioglu, A. Iqbal, J. Swales, R. Owen, M. Higgins, B. Kramer, and INHANCE Study Investigators (2010). Once-daily bronchodilators for chronic obstructive pulmonary disease: Indacaterol versus tiotropium. *Am. J. Respir. Crit. Care Med. 182*(2), 155–162.

Dunnett, C. W. (1955). A multiple comparison procedure for comparing several treatments with a control. *Journal of the American Statistical Association 50*(272), 1096–1121.

Hampson, L. V. and C. Jennison (2015). Optimizing the data combination rule for seamless phase II/III clinical trials. *Statistics in Medicine 34*(1), 39–58.

Huang, X., J. Ning, Y. Li, E. Estey, J. P. Issa, and D. A. Berry (2009). Using short-term response information to facilitate adaptive randomization for survival clinical trials. *Statistics in Medicine 28*(12), 1680–1689.

Iacono, A. T., B. A. Johnson, W. F. Grgurich, J. G. Youssef, T. E. Corcoran, D. A. Seiler, J. H. Dauber, G. C. Smaldone, A. Zeevi, S. A. Yousem, J. J. Fung, G. J. Burckart, K. R. McCurry, and B. P. Griffith (2006). A randomized trial of inhaled cyclosporine in lung-transplant recipients. *New England Journal of Medicine 354*, 141–150.

Inoue, L. Y., P. F. Thall, and D. A. Berry (2002). Seamlessly expanding a randomized phase II trial to phase III. *Biometrics 58*(4), 823–831.

Jakob, S. M., E. Ruokonen, R. M. Grounds, T. Sarapohja, C. Garratt, S. J. Pocock, J. R. Bratty, and J. Takala (2012). Dexmedetomidine vs midazolam or propofol for sedation during prolonged mechanical ventilation: Two randomized controlled trials. *Journal of the American Medical Association 307*, 1151–1160.

Jennison, C. and B. W. Turnbull (2007). Adaptive seamless designs: Selection and prospective testing of hypotheses. *Journal of Biopharmaceutical Statistics 17*(6), 1135–1161.

Kahan, B. C., V. Jairath, C. J. Doré, and T. P. Morris (2014). The risks and rewards of covariate adjustment in randomized trials: an assessment of 12 outcomes from 8 studies. *Trials 15*(1), 139.

Koenig, F., W. Brannath, F. Bretz, and M. Posch (2008). Adaptive Dunnett tests for treatment selection. *Statistics in Medicine 27*(10), 1612–1625.

Krueger, G. G., R. G. Langley, C. Leonardi, N. Yeilding, C. Guzzo, Y. Wang, L. T. Dooley, and M. Lebwohl (2007). A human interleukin-12/23 monoclonal antibody for the treatment of psoriasis. *New England Journal of Medicine 356*, 580–592.

Lehmacher, W. and G. Wassmer (1999). Adaptive sample size calculations in group sequential trials. *Biometrics 55*(4), 1286–1290.

Liu, Q., M. A. Proschan, and G. W. Pledger (2002). A unified theory of two-stage adaptive designs. *Journal of the American Statistical Association 97*(460), 1034–1041.

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

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

Magirr, D., T. Jaki, and J. Whitehead (2012). A generalized Dunnett test for multi-arm multi-stage clinical studies with treatment selection. *Biometrika 99*(2), 494–501.

Marcus, R., P. Eric, and K. R. Gabriel (1976). On closed testing procedures with special reference to ordered analysis of variance. *Biometrika 63*(3), 655–660.

Müller, H. H. and H. Schäfer (2001). Adaptive group sequential designs for clinical trials: Combining the advantages of adaptive and of classical group sequential approaches. *Biometrics 57*(3), 886–891.

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

Posch, M., F. Koenig, M. Branson, W. Brannath, C. Dunger-Baldauf, and P. Bauer (2005). Testing and estimation in flexible group sequential designs with adaptive treatment selection. *Statistics in Medicine 24*(24), 3697–3714.

Prowell, T. M., M. R. Theoret, and R. Pazdur (2016). Seamless oncology-drug development. *New England Journal of Medicine 374*(21), 2001–2003.

Sarkar, S. K. and C. K. Chang (1997). The Simes method for multiple hypothesis testing with positively dependent test statistics. *Journal of the American Statistical Association 92*(440), 1601–1608.

Schmidli, H., F. Bretz, A. Racine, and W. Maurer (2006). Confirmatory seamless phase II/III clinical trials with hypotheses selection at interim: Applications and practical considerations. *Biometrical Journal 48*(4), 635–643.

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

Simes, R. J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika 73*(3), 751–754.

Stallard, N. and T. Friede (2008). A group-sequential design for clinical trials with treatment selection. *Statistics in Medicine 27*(29), 6209–6227.

Sverdlov, O. (2015). *Modern Adaptive Randomized Clinical Trials: Statistical and Practical Aspects*, Volume 81. CRC Press.

Thall, P. F., R. Simon, and S. S. Ellenberg (1988). Two-stage selection and testing designs for comparative clinical trials. *Biometrika 75*(2), 303–310.

US Food and Drug Administration (2006). Innovation or stagnation: Critical path opportunities report and list.

US Food and Drug Administration (2018). Expansion cohorts: Use in first-in-human clinical trials to expedite development of oncology drugs and biologics guidance for industry.

Yuan, Y., X. Huang, and S. Liu (2011). A Bayesian response-adaptive covariate-balanced randomization design with application to a leukemia clinical trial. *Statistics in Medicine 30*(11), 1218–1229.

Zang, Y. and J. J. Lee (2014). Adaptive clinical trial designs in oncology. *Chinese Clinical Oncology 3*(4).

Zhu, H. and F. Hu (2019). Sequential monitoring of covariate-adaptive randomized clinical trials. *Statistica Sinica 29*, 265–282.