Contemporary Clinical Trials 87 (2019) 105874

Contents lists available at ScienceDirect

Contemporary Clinical Trials

journal homepage: www.elsevier.com/locate/conclintrial

Sequential monitoring of covariate adaptive randomized clinical trials with sample size re-estimation

Jun Yu, Dejian Lai*

Department of Biostatistics ad Data Science, University of Texas Health Science Center at Houston, Houston, TX, USA

ARTICLE INFO

Keywords:
Adaptive design
Interim analysis
Power
Type I error

ABSTRACT

Once scientific questions are determined, other design features of clinical trials including increasing the power while controlling the type I error rate, planning interim analysis, and achieving treatment balance among subgroups will be either required or preferred. We propose to sequentially monitor the covariate adaptive randomization (CAR) procedures with sample size re-estimation (SSR) to satisfy a variety of design objectives of clinical trials. However, each of the three adaptive designs (sequential monitoring, CAR, and SSR) poses a challenge to the control of the type I error rate. In this paper, we investigated how to utilize the advantages of the three adaptive methods and control the type I error rate. We proved that the asymptotic joint distribution of the sequential statistics follows the asymptotic canonical joint distribution defined in Jennison and Turnbull [14]. Besides, numerical studies demonstrated that our methods could control the type I error rate, increase the power, and lead to much-improved treatment balance across subgroups.

1. Introduction

Clinical trials usually involve competitive objectives such as increasing the power while controlling the type I error rate, planning interim analysis, and adjusting the sample size. In this paper, we propose to sequentially monitor the covariate adaptive randomized clinical trials with sample size re-estimation (SSR) to satisfy different objectives of clinical trials. We will study the feasibility and advantages of this procedure.

Sequential monitoring can prevent patients from being exposed to unsafe or ineffective treatment regimens, ensure that the protocol is complied and save cost, time, and available patient resources [13]. Sequential monitoring stemmed from the sequential probability ratio test [26], and the most influential papers include Pocock [19], O'Brien and Fleming [18], and Lan and DeMets [15].

Sample size re-estimation is also important in designing a clinical trial because the originally planned sample size is often calculated based on prior studies that used different medical practices, participating populations, etc. Accordingly, the power may be unexpectedly low due to a poorly calculated sample size. Cui et al. [3] and Denne [4] specially studied how to implement SSR for group sequential tests.

It is well accepted that the balance of treatment allocation among subgroups defined by covariates is critical to assess the treatment effects in clinical trials properly. Covariate adaptive randomization (CAR) procedures sequentially assign the patients based on previous treatment assignments and covariates, and the current covariate profile to achieve this aim and to increase the credibility of a trial [22]. Stratified permuted block (SPB) randomization [28] and the Pocock and Simon's design [20] are the two most popular CAR procedures. Other CAR designs and clinical trials using CAR include Baldi Antognini and Zagoraiou [1], Iacono et al. [12], Barnes et al. [2], and Hu and Hu [11].

It is desirable to study how to benefit from the advantages of the three adaptive approaches (group sequential monitoring, CAR, and SSR) in one clinical trial. However, each of the three methods poses a challenge to the control of the type I error rate that is the critical issue of confirmatory clinical trials. Sequential monitoring leads to correlated sequential statistics and inflates the type I error rate due to multiple interim looks; CAR results in complicated dependence among covariates, treatment assignments, and responses; and SSR not only changes the maximum information but also brings about dependence among the observed data. Therefore, the derivation of the joint distribution of the sequential statistics and the control the type I error rate are complicated. In this paper, we overcame these difficulties, proposed approaches to control the type I error rate for sequential monitoring of CAR with SSR and studied its asymptotic and finite-sample properties.

In Section 2, we introduce the general framework, notation, our proposed methods, and theoretical findings for sequential monitoring of CAR with SSR. In Section 3, we focus on two special cases. In Section 4,

* Corresponding author at: 1200 Pressler Street, RAS W1008, Houston, TX 77030, USA.
E-mail address: dejian.lai@uth.tmc.edu (D. Lai).

https://doi.org/10.1016/j.cct.2019.105874
Received 28 August 2019; Received in revised form 16 October 2019; Accepted 24 October 2019
Available online 31 October 2019
1551-7144/

we report results from numerical studies. Conclusions are in Section 5, and the proof is in the Supplementary materials.

2. Sequential monitoring of CAR with SSR

2.1. Framework

In this paper, we focus on a common situation in real clinical trials that raises lots of concerns: only some of the randomization covariates are included in the data analysis, or none of these covariates are included (*t*-test). For example, Lai et al. [14] used permuted block randomization stratified on gender in a clinical trial to study the effects of music on the kangaroos' maternal anxiety since female kangaroo infants seem to be more likely to survive than male infants. However, a *t*-test was performed in the data analysis. There are many practical reasons for the clinical trialists' reluctance to use all the randomization covariates: (i) it is often difficult to explain the practical meaning of the effects of certain covariates such as the investigation sites; (ii) it leads to the theoretical difficulties to include a large number of covariates in the model; (iii) and it is also difficult to justify the model specification for some covariates to be included in the model (e.g., whether interaction terms should be included). Shao et al. [23], Ma et al. [17], and Zhu and Hu [30] studied the statistical inference for this problem, but they did not consider SSR in their works.

Consider a two-arm randomized controlled clinical trial with originally planned n subjects to be sequentially allocated by CAR procedures. Let T_i ($i = 1, \ldots, n$) be the treatment assignment ($T_i = 1$ if treatment 1; $T_i = 0$ if treatment 2). Assume that the covariates ($X_1, \ldots, X_p$) and ($V_1, \ldots, V_q$) are used to implement CAR, but only ($X_1, \ldots, X_p$) are included in the data analysis. For simplicity, we only consider one-dimensional covariates, but it is easy to generalize the results to multi-dimensional covariates. Assume that all the covariates are independent and centered at 0 without loss of generality, i.e., $E(X_{ik}) = 0$, $E(V_{ij}) = 0$, $i = 1, \ldots, n$, $k = 1, \ldots, p$, $j = 1, \ldots, q$. Besides, the errors are assumed to be independent with the covariates. This general framework includes two special cases. First, if only ($X_1, \ldots, X_p$) are used to implement CAR, we are using all the randomization covariates in the data analysis. Second, if only ($V_1, \ldots, V_q$) are used to implement CAR and no covariates are used in the data analysis, we are studying *t*-test. These two special cases will be emphasized in the next section.

Assume that the ith subject's response Y_i follows the linear model:

$$Y_i = \mu_1 T_i + \mu_2 (1 - T_i) + X_{i1}\beta_1 + \ldots + X_{ip}\beta_p + V_{i1}\gamma_1 + \ldots + V_{iq}\gamma_q + \epsilon_i \quad (1)$$

where μ_1 and μ_2 are the treatment effects for the treatments 1 and 2, respectively, $(\beta_1, \ldots, \beta_p)$ and $(\gamma_1, \ldots, \gamma_q)$ are unknown parameters for covariate effects, and the ϵ_i are independent errors with mean 0 and variance σ^2. Here, we do not require that the errors follow normal distribution. Write $\boldsymbol{\mu} = (\mu_1, \mu_2)^T$, $\boldsymbol{\eta} = (\mu_1, \mu_2, \beta_1, \ldots, \beta_p)^T$, $\boldsymbol{T}(n) = (T_1, \ldots, T_n)^T$, $\boldsymbol{Y}(n) = (Y_1, \ldots, Y_n)^T$, $\epsilon(n) = (\epsilon_1, \ldots, \epsilon_n)^T$ and

$$\boldsymbol{X}(n) = \begin{bmatrix} T_1 & 1 - T_1 & X_{11} & \ldots & X_{1p} \\ T_2 & 1 - T_2 & X_{21} & \ldots & X_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ T_n & 1 - T_n & X_{n1} & \ldots & X_{np} \end{bmatrix}$$

CAR designs are usually applied with discrete covariates. When implementing CAR using continuous covariates, we first discretize these continuous covariates and apply CAR designs concerning the discretized covariates. Specifically, let

$$\widetilde{X}_j = \begin{cases} X_j & \text{if } j \notin C \\ d_j(X_j) & \text{if } j \in C \end{cases}$$

and

$$\widetilde{V}_j = \begin{cases} V_j & \text{if } j \notin C^* \\ d_j^*(V_j) & \text{if } j \in C^* \end{cases}$$

where $C = \{l : \text{index of continuous covariates among } X_l, l = 1, \ldots, p\}$, $C^* = \{l : \text{index of continuous covariates among } V_l, l = 1, \ldots, q\}$, and $d_j(\cdot)$ and $d_j^*(\cdot)$ are discrete functions. Usually, the cut-off points to discretize continuous covariates depend on the scientific question, previous experience, and the physician's point of view about the disease. For example, in the guidelines for the management of Aneurysmal Subarachnoid Hemorrhage (2012), Aneurysms in the anterior circulation is believed to be more likely to break in patients < 55 years of age. Therefore, 55 will be a natural cut-off point for age in this case. Note that our theorems in this paper are applicable when we use original continuous covariates in the data analysis instead of the discretized covariates.

We perform the following hypothesis testing to compare two treatments in clinical trials:

$$H_0: \mu_1 = \mu_2 \text{ versus } \mu_1 \neq \mu_2 \quad (2)$$

Let $\lfloor \cdot \rfloor$ denote the floor function and $t = N/n$ be the *information time* when N is the number of enrolled patients. Then the natural sequential statistics to test the hypothesis (2) at time point $t \in (0, 1]$ by fitting the linear model without $(V_1, \ldots, V_q)$ is

$$Z_t = \frac{L\hat{\boldsymbol{\eta}}(t)}{\sqrt{\hat{\sigma}(t)^2 L(\boldsymbol{X}(\lfloor nt \rfloor)^T \boldsymbol{X}(\lfloor nt \rfloor))^{-1} L^T}} \quad (3)$$

where $L = (1, -1, 0, \ldots, 0)$, $\hat{\boldsymbol{\eta}}(t) = (\boldsymbol{X}(\lfloor nt \rfloor)^T \boldsymbol{X}(\lfloor nt \rfloor))^{-1} \boldsymbol{X}(\lfloor nt \rfloor)^T \boldsymbol{Y}(\lfloor nt \rfloor)$,ss

$$\hat{\sigma}(t)^2 = [\boldsymbol{Y}(\lfloor nt \rfloor) - \boldsymbol{X}(\lfloor nt \rfloor)\hat{\boldsymbol{\eta}}(t)]^T [\boldsymbol{Y}(\lfloor nt \rfloor) - \boldsymbol{X}(\lfloor nt \rfloor)\hat{\boldsymbol{\eta}}(t)] / (\lfloor nt \rfloor - p - 2).$$

We have introduced a general framework for sequential monitoring of covariate adaptive randomized clinical trials. Note that the above procedure and sequential statistics will not cause a problem if complete randomization is used [30]. Next, we will introduce how to implement the SSR and control the type I error rate in clinical trials with CAR.

2.2. Incorporation of sample size re-estimation

In this paper, we only allow an increase of sample size as recommended by [6]. Suppose we have K interim analyses at information time points $t_1, \ldots, t_L, \ldots, t_K$, and we implement SSR at the end of the Lth interim analysis ($L < K$) based on the observed data using the method in Cui et al. [3]. Let

$$Z_t^{adj} = \frac{L\hat{\boldsymbol{\eta}}(t)}{\hat{\varepsilon}(t)\sqrt{\hat{\sigma}(t)^2 L(\boldsymbol{X}(\lfloor nt \rfloor)^T \boldsymbol{X}(\lfloor nt \rfloor))^{-1} L^T}} \quad (4)$$

where $\hat{\varepsilon}(t)^2$ is any consistent estimator of

$$\frac{\sum_{j \in C^*} \gamma_j^2 \sigma_{\delta j}^2 + \sigma^2}{\sigma^2 + \sum_{j=1}^{q} Var(V_j \gamma_j^T)} \quad (5)$$

$\sigma_{\delta j}{}^2 = E[Var(\delta_j | d_j^*(V_j))]$, and $\delta_j = V_j - E(V_j | d_j^*(V_j))$. We first calculate the conditional power, CP_L, based on observed data for originally planned sample size n. If CP_L is not less than the desirable level of cp, then no SSR will be implemented. Otherwise, search n^* that satisfies $CP_L = cp$. Next, we increase the original sample size at stages $k \geq L + 1$ by a multiplier of $b = \min(b^*, b_{\max})$, where $b_{\max}$ is a prespecified maximum sample size factor, and $b^* = (n^* - N)/(n - N)$.

Then we can use the following new sequential statistics to perform sequential monitoring

$$U_t = \begin{cases} Z_t^{adj}, & \text{if } t \le t_L; \\ w_t^{1/2} \times Z_{t_L}^{adj} + (1-w_t)^{1/2} \times \\ \quad \{[B(b(t-t_L)+t_L) - B(t_L)]/[b(t-t_L)]^{1/2}\}, & \text{if } t > t_L, \end{cases} \quad (6)$$

where $w_t = t_L/t$, $B(t) = \sqrt{t} Z_t^{adj}$.

2.3. Asymptotic results

For the proposed procedure, we observe a variety of random processes such as $(T_1, \ldots, T_{\lfloor nt \rfloor})$, $(Y_1, \ldots, Y_{\lfloor nt \rfloor})$, and those for covariates and sequential parameter estimators. When CAR is used, these random processes are related to each other in a complicated manner. For example, $(T_1, \ldots, T_{\lfloor nt \rfloor})$ are dependent on the covariates; any observed response depends on all the previous treatment assignments, covariates, and responses. All these complex relationships lead to difficulties in deriving the joint distributions of sequential test statistics.

We need the following notation to formulate the main theorem. Let $\boldsymbol{W}_i = (x_{i1}^{c_1}, \ldots, x_{ip}^{c_p}, v_{i1}^{c_1^*}, \ldots, v_{iq}^{c_q^*})$ represents the ith subject's covariate profile if $\widetilde{X}_{ik}$ is at level $x_{ik}^{c_k}$, $k = 1, \ldots, p$ and $\widetilde{V}_{ij}$ is at level $v_{ij}^{c_j^*}$, $j = 1, \ldots, q$. Let DIF_n be the overall difference in patient numbers between two treatments after n patients have been enrolled in the trial; similarly, let $DIF_n^X(k; c_k)$ be the marginal difference concerning the level $x_k^{c_k}$ of covariate $\widetilde{X}_k$ and $DIF_n^V(j; c_j^*)$ be the marginal difference concerning the level $v_j^{c_j^*}$ of covariate $\widetilde{V}_j$; let $DIF_n(c_1, \ldots, c_p, c_1^*, \ldots, c_q^*)$ be the difference in patient numbers in the stratum containing the subjects with covariates $(x_1^{c_1}, \ldots, x_p^{c_p}, v_1^{c_1^*}, \ldots, v_q^{c_q^*})$.

Theorem 1. Let $B_t^U = \sqrt{t} U_t$. Assume the CAR design satisfies $DIF_n = O_P(1)$, $DIF_n^X(k; c_k) = O_P(1)$, $k = 1, \ldots, p$, and $DIF_n^V(j; c_j^*) = O_P(1)$, $j = 1, \ldots, q$. Then under H_0, B_t^U is asymptotically a standard Brownian motion in distribution, and the sequence of test statistics $\{(U_{t_1}, \ldots, U_{t_k}), 0 \le t_1 \le t_2 \le \ldots \le t_K \le 1\}$ satisfies.

(ii) $\{U_{t_1}, \ldots, U_{t_k}\}$ *follows multivariate normal distribution*;
(iii) $EU_{t_i} = 0$;
(iiii) $Cov(U_{t_i}, U_{t_j}) = \sqrt{\lfloor nt_i \rfloor / \lfloor nt_j \rfloor}$, $0 \le t_i \le t_j \le 1$.

Theorem 1 reveals one of the most fundamental properties for the proposed method. That is, the asymptotic joint distribution of the sequential statistics follows the asymptotic canonical joint distribution defined in Jennison and Turnbull [13]. Moreover, the conditions required in this theorem hold for many commonly used CAR procedures such as the stratified permuted block design and the Pocock and Simon's design [20, 17].

From Theorem 1, we can easily see that the type I error rates would be conservative if we used Z_t in (3) instead of Z_t^{adj} in (4) to calculate U_t when fitting a linear model without $(V_1, \ldots, V_q)$ since

$$\frac{\sum_{j \in C^*} \gamma_j^2 \sigma_{\delta j}^2 + \sigma^2}{\sigma^2 + \sum_{j=1}^{q} Var(V_j \gamma_j^T)}$$

is always less than 1. Accordingly, the power will be adversely affected, and the required sample size needs to increase, which attenuates the advantages of sequential monitoring.

Because the sequential statistics converge to the asymptotic canonical joint distribution defined in Jennison and Turnbull [13], all the approaches based on this distribution in that book and in the literature such as the equivalence test, spending functions, stochastic curtailment, and repeated confidence intervals can be used to either control the type I error rate or to offer valuable information for the Data and Safety Monitoring Board to make decision about whether to stop the trial.

In the numerical studies, we use the following spending function [15,21] to approximates the O'Brien-Fleming boundaries [18]

$$\alpha_1(t) = 2\{1 - \Phi(z_{\alpha/2}/t^{1/2})\}.$$

This method does not require the predetermined number of interim analysis and equally spaced analysis. The O'Brien-Fleming-like function spends little of the type I error at the early interim analysis, and the last critical value is very close to that for clinical trials without sequential monitoring. All these features are preferred by the Data and Safety Monitoring Boards (DSMB). Note that the critical values are corresponding to the process U_t.

3. Two special cases

As mentioned before, we consider two special cases.

Special Case 1: If $(\gamma_1, \ldots, \gamma_q) = (0, \ldots, 0)$ in model (1) and only $(X_1, \ldots, X_p)$ are used to implement CAR, the framework and notation in Section 2 are applicable for the scenario where all the randomization covariates are used in data analysis. Specifically, we assume that the ith subject's response Y_i follows the linear model:

$$Y_i = \mu_1 T_i + \mu_2 (1 - T_i) + X_{i1}\beta_1 + \ldots + X_{ip}\beta_p + \epsilon_i \quad (7)$$

The other design and notation are the same as in Section 2.

We propose the same SSR procedure as in Section 2, with the following sequential statistics

$$U_t^{full} = \begin{cases} Z_t, & \text{if } t \le t_L; \\ w_t^{1/2} \times Z_{t_L} + (1-w_t)^{1/2} \times \\ \quad \{[B(b(t-t_L)+t_L) - B(t_L)]/[b(t-t_L)]^{1/2}\}, & \text{if } t > t_L, \end{cases} \quad (8)$$

where $w_t = t_L/t$, $B(t) = \sqrt{t} Z_t$, and Z_t was defined in (3). We have the following theorem for this special case.

Theorem 2. Let $B_t^{full} = \sqrt{t} U_t^{full}$. Assume the CAR design satisfies $DIF_n = O_P(1)$ and $DIF_n^X(k; c_k) = O_P(1)$, $k = 1, \ldots, p$. Then under H_0, B_t^{full} is asymptotically a standard Brownian motion in distribution, and the sequential statistics $\{(U_{t_1}^{full}, \ldots, U_{t_k}^{full}), 0 \le t_1 \le t_2 \le \ldots \le t_K \le 1\}$ satisfies.

(ii) $\{U_{t_1}^{full}, \ldots, U_{t_k}^{full}\}$ *follows multivariate normal distribution*;
(iii) $EU_{t_i}^{full} = 0$;
(iiii) $Cov(U_{t_i}^{full}, U_{t_j}^{full}) = \sqrt{\lfloor nt_i \rfloor / \lfloor nt_j \rfloor}$, $0 \le t_i \le t_j \le 1$.

Special Case 2: The other special case is that none of the randomization covariates are used in the data analysis, and the reasons for the popularity of this approach were given in the Introduction. In this case, we are performing the t-test or equivalently fitting the following model:

$$Y_i = \mu_1 T_i + \mu_2 (1 - T_i) + \epsilon_i, \; i = 1, \ldots, n. \quad (9)$$

Here, the CAR is implemented concerning $(V_1, \ldots, V_q)$, and the responses follow:

$$Y_i = \mu_1 T_i + \mu_2 (1 - T_i) + V_{i1}\gamma_1 + \ldots + V_{iq}\gamma_q + \epsilon_i, \; i = 1, \ldots, n \quad (10)$$

Let $Q = (1, -1)$ and

$$\boldsymbol{Tr}(n) = \begin{bmatrix} T_1 & 1 - T_1 \\ T_2 & 1 - T_2 \\ \vdots & \vdots \\ T_n & 1 - T_n \end{bmatrix}$$

We propose the same SSR procedure as in Section 2, with the following sequential statistics

$$U_t^{ttest} = \begin{cases} Z_t^{ttest}, & \text{if } t \le t_L; \\ w_t^{1/2} \times Z_{t_L}^{ttest} + (1-w_t)^{1/2} \times \\ \quad \{[B(b(t-t_L)+t_L) - B(t_L)]/[b(t-t_L)]^{1/2}\}, & \text{if } t > t_L, \end{cases} \quad (11)$$

where $w_t = t_L/t$, $B(t) = \sqrt{t} Z_t^{ttest}$, and Z_t^{ttest} was defined as

$$Z_t^{ttest} = \frac{Q\hat{\boldsymbol{\mu}}(t)}{\hat{\epsilon}(t)\sqrt{\hat{\sigma}(t)^2 Q(\boldsymbol{Tr}(\lfloor nt \rfloor)^T \boldsymbol{Tr}(\lfloor nt \rfloor))^{-1} Q^T}} \quad (12)$$

Table 1
Performance of different designs under H_0 when all the randomization covariates are included in the data analysis.

Both covariates are discrete												
(β_0, p_1, p_2)	Design	α	$\widehat{\beta}_T$	DIF_n	DIF_{11}	DIF_{10}	DIF_{01}	DIF_{00}	$DIF_{1\cdot}$	$DIF_{0\cdot}$	$DIF_{\cdot 1}$	$DIF_{\cdot 0}$
(0.5,0.5,0.5)	SPB	0.049	0.000 (0.080)	1.32 (1.27)	0.67 (0.62)	0.66 (0.63)	0.67 (0.62)	0.67 (0.62)				
(0.5,0.5,0.5)	PS	0.051	0.000 (0.080)	1.70 (1.70)	5.44 (4.16)	5.44 (4.17)	5.46 (4.16)	5.44 (4.15)	1.54 (1.42)	1.52 (1.38)	1.53 (1.38)	1.54 (1.40)
(0.5,0.5,0.5)	CR	0.051	0.001 (0.080)	20.7 (15.7)	10.4 (7.91)	10.5 (7.89)	10.6 (8.11)	10.5 (8.02)	14.5 (11.1)	14.8 (11.4)	14.8 (11.4)	14.7 (11.2)
(2,0.4,0.6)	SPB	0.050	0.000 (0.079)	1.33 (1.27)	0.66 (0.62)	0.68 (0.63)	0.67 (0.62)	0.66 (0.62)				
(2,0.4,0.6)	PS	0.051	0.000 (0.081)	1.68 (1.66)	5.22 (4.00)	5.19 (3.96)	5.26 (4.03)	5.22 (4.00)	1.52 (1.41)	1.55 (1.38)	1.51 (1.37)	1.50 (1.38)
(2,0.4,0.6)	CR	0.050	0.000 (0.079)	20.8 (16.2)	10.3 (7.84)	8.42 (6.40)	12.6 (9.54)	10.3 (7.83)	13.3 (10.1)	16.4 (12.4)	16.2 (12.4)	13.3 (10.1)
Both covariates are continuous												
(β_0, q_1, q_2)	Design	α	$\widehat{\beta}_T$	DIF_n	DIF_{11}	DIF_{10}	DIF_{01}	DIF_{00}	$DIF_{1\cdot}$	$DIF_{0\cdot}$	$DIF_{\cdot 1}$	$DIF_{\cdot 0}$
(0.5,0.5,0.5)	SPB	0.055	0.000 (0.081)	1.32 (1.28)	0.68 (0.62)	0.67 (0.62)	0.66 (0.62)	0.67 (0.63)				
(0.5,0.5,0.5)	PS	0.049	−0.001 (0.080)	1.69 (1.68)	5.43 (4.14)	5.43 (4.15)	5.43 (4.12)	5.42 (4.13)	1.53 (1.38)	1.55 (1.41)	1.53 (1.39)	1.53 (1.39)
(0.5,0.5,0.5)	CR	0.055	0.000 (0.082)	20.8 (16.0)	10.5 (8.04)	10.5 (8.01)	10.4 (7.94)	10.4 (7.99)	14.8 (11.4)	14.7 (11.3)	14.8 (11.4)	14.8 (11.4)
(2,0.4,0.6)	SPB	0.054	0.000 (0.080)	1.33 (1.27)	0.66 (0.62)	0.67 (0.63)	0.67 (0.62)	0.67 (0.62)				
(2,0.4,0.6)	PS	0.052	0.001 (0.081)	1.71 (1.69)	5.17 (3.95)	5.09 (3.91)	5.16 (3.97)	5.12 (3.94)	1.52 (1.39)	1.52 (1.40)	1.53 (1.38)	1.54 (1.40)
(2,0.4,0.6)	CR	0.057	0.000 (0.082)	21.2 (16.0)	10.0 (7.79)	8.42 (6.34)	12.7 (9.67)	10.2 (7.84)	13.2 (10.0)	16.4 (12.6)	16.2 (12.2)	13.2 (10.2)

Note: α: type I error rate; $\widehat{\beta}_T$: estimator of β_T; DIF_n: overall difference in patient numbers between the two treatments; DIF_{gh} for $X_1 = g$ and $X_2 = h$, g, $h = 0, 1$: the differences of patient numbers between the two treatments in the four stratum; $DIF_{1\cdot}$: marginal imbalance for $X_1 = 1$; $DIF_{0\cdot}$: marginal imbalance for $X_1 = 0$; $DIF_{\cdot 1}$: marginal imbalance for $X_2 = 1$; $DIF_{\cdot 0}$: marginal imbalance for $X_2 = 0$.

where $\widehat{\boldsymbol{\mu}}(t) = (\boldsymbol{Tr}(\lfloor nt \rfloor)^T \boldsymbol{Tr}(\lfloor nt \rfloor))^{-1} \boldsymbol{Tr}(\lfloor nt \rfloor)^T \boldsymbol{Y}(\lfloor nt \rfloor)$,

$$\widehat{\sigma}(t)^2 = [\boldsymbol{Y}(\lfloor nt \rfloor) - \boldsymbol{Tr}(\lfloor nt \rfloor)\widehat{\boldsymbol{\mu}}(t)]^T [\boldsymbol{Y}(\lfloor nt \rfloor) - \boldsymbol{Tr}(\lfloor nt \rfloor)\widehat{\boldsymbol{\mu}}(t)]/(\lfloor nt \rfloor - 2),$$

and $\widehat{\varepsilon}(t)^2$ is a consistent estimator of

$$\frac{\sum_{j \in C^*} \gamma_j^2 \sigma_{\delta j}^2 + \sigma^2}{\sigma^2 + \sum_{j=1}^{q} Var(V_j \gamma_j^T)}$$

Then we have the following theorem for this special case.

Theorem 3. Let $B_t^{ttest} = \sqrt{t}\, U_t^{ttest}$. Assume the CAR design satisfies $DIF_n = O_P(1)$ and $DIF_n{}^V(j; c_j^*) = O_P(1)$, $j = 1, \ldots, q$. Then under H_0, B_t^{ttest} is asymptotically a standard Brownian motion in distribution, and the sequence of test statistics $\{(U_{t_1}^{ttest}, \ldots, U_{t_k}^{ttest}), 0 \le t_1 \le t_2 \le \ldots \le t_K \le 1\}$ satisfies.

(ii) $\{U_{t_1}^{ttest}, \ldots, U_{t_k}^{ttest}\}$ *follows multivariate normal distribution*;
(iii) $EU_{t_i}^{ttest} = 0$;
(iiii) $Cov\left(U_{t_i}^{ttest}, U_{t_j}^{ttest}\right) = \sqrt{\lfloor nt_i \rfloor / \lfloor nt_j \rfloor}$, $0 \le t_i \le t_j \le 1$.

As in Theorem 1, Theorem 3 also explicitly shows the problems in clinical trials with CAR if our analysis is based on *t*-test and the theory of complete randomization. In the next section, we will numerically demonstrate this problem and discuss our proposed methods to control the type I error rate, increase the power, and improve the treatment balances among subgroups.

4. Numerical studies

In this section, we study the finite-sample properties of our proposed procedure. We will focus on the two special cases in Section 3. We also performed numerical studies for the scenario where part of the randomization covariates was included in the data analysis, but the conclusions were similar to the special case that no covariates were included in the data analysis. Therefore, we skip the report for that scenario in this paper.

4.1. Numerical results for special case 1

Suppose originally planned 500 patients sequentially enter a clinical trial, and the responses follow

$$Y_i = \mu_1 T_i + \mu_2 (1 - T_i) + X_{i1}\beta_1 + X_{i2}\beta_2 + \epsilon_i,\ i = 1, \ldots, 500 \quad (13)$$

where $(\mu_1, \mu_2, \beta_1, \beta_2)$ are unknown parameters, and ε_i are independent errors from the normal distribution $N(0, 1)$. We compare the stratified permuted block randomization (SPB), Pocock and Simon's procedure (PS) and complete randomization (CR). The two CAR designs were applied concerning both X_1 and X_2. We consider two cases: 1) both X_1 and X_2 are binary covariates with success rates of p_1 and p_2, respectively; 2) both X_1 and X_2 follow the standard normal distribution. When the CAR procedures are implemented with continuous X_j, $j = 1, 2$, we discretize them in the following way:

$$\bar{x} = \begin{cases} 1 & \text{if } x < z_{q_j} \\ 0 & \text{if } x \ge z_{q_j} \end{cases},$$

where z_{q_j} is the q_j-quantile of the standard normal distribution. However, the original continuous covariates will be included in the data analysis.

The sequential data analysis is based on model (13). Equivalently, it can be written as

$$Y_i = \beta_0 + \beta_T T_i + X_{i1}\beta_1 + X_{i2}\beta_2 + \epsilon_i,\ i = 1, \ldots, 500 \quad (14)$$

That is, all the randomization covariates are used in the data analysis. We plan to perform sequential monitoring at information time $t_1 = 0.2$, $t_2 = 0.5$, and $t_3 = 1$, and the corresponding O'Brien-Fleming-like boundaries are $C_1 = 4.877$, $C_2 = 2.963$, $C_3 = 1.969$. We implement SSR if the trial is determined to continue after the second interim analysis. The cap of the sample size at stage 3 is 500 and $b_{\max} = 2$.

We first study the performance of different designs under H_0. We set $\beta_1 = \beta_2 = 1$ and vary other parameters. In Table 1, we report the type I error rate (α) and $\widehat{\beta}_T$ as a representative of the parameter estimators. We study the overall imbalance and stratum-level imbalance for SPB. In addition to those imbalances, we also report the marginal imbalance for Pocock and Simon's procedure since it aims to control the marginal imbalance. All these features of our proposed methods are compared with those of complete randomization. The standard deviations are given in parentheses. All the results are based on 10,000 replications.

From Table 1, we can see that when all the randomization covariates are included in the data analysis, both our methods and complete randomization can control the type I error rate well and accurately estimate the unknown parameters. SPB can lead to significantly better overall balance and stratum-level balance than complete randomization, and it also outperforms the Pocock and Simon's design. Compared to complete randomization, Pocock and Simon's design also leads to

Table 2
Type I error rate of different designs under H_0 when unadjusted t-test is used.

(β_0, p_1, p_2)	X_1, X_2	Design	α	(β_0, q_1, q_2)	X_1, X_2	Design	α
(0.5,0.5,0.5)	Discrete	SPB	0.013	(0.5,0.5,0.5)	Continuous	SPB	0.006
(0.5,0.5,0.5)	Discrete	PS	0.015	(0.5,0.5,0.5)	Continuous	PS	0.007
(0.5,0.4,0.6)	Discrete	SPB	0.013	(0.5,0.4,0.6)	Continuous	SPB	0.008
(0.5,0.4,0.6)	Discrete	PS	0.014	(0.5,0.4,0.6)	Continuous	PS	0.010
(2,0.5,0.5)	Discrete	SPB	0.010	(2,0.5,0.5)	Continuous	SPB	0.007
(2,0.5,0.5)	Discrete	PS	0.016	(2,0.5,0.5)	Continuous	PS	0.006
(2,0.4,0.6)	Discrete	SPB	0.011	(2,0.4,0.6)	Continuous	SPB	0.006
(2,0.4,0.6)	Discrete	PS	0.012	(2,0.4,0.6)	Continuous	PS	0.007

much better overall, stratum-level, and marginal balance. The control of the balances at different levels is the main advantages of CAR design for this special case, while CAR design can improve the power marginally. We omit the results for power of Special Case 1 and will emphasize the advantages of our methods in terms of power in the next section.

4.2. Numerical results for special case 2

Although including all the randomization covariates in the data analysis has been recommended, clinical trial practitioners often prefer t-test without using any covariates in the data analysis for the reasons mentioned in the Introduction. We first detect the problems of this approach and report the results in Table 2. We follow the same simulation setting in Section 4.1 except that we use

$$U_t^{tt} = \begin{cases} Z_t^{tt}, & \text{if } t \le t_L; \\ w_t^{1/2} \times Z_{t_L}^{tt} + (1-w_t)^{1/2} \times \\ \quad \{[B(b(t-t_L)+t_L) - B(t_L)]/[b(t-t_L)]^{1/2}\}, & \text{if } t > t_L \end{cases} \tag{15}$$

where $w_t = t_L/t$, $B(t) = \sqrt{t} Z_t^{tt}$, and Z_t^{tt} was defined as

$$Z_t^{tt} = \frac{Q\hat{\boldsymbol{\mu}}(t)}{\sqrt{\hat{\sigma}(t)^2 Q(\boldsymbol{Tr}(\lfloor nt \rfloor)^T \boldsymbol{Tr}(\lfloor nt \rfloor))^{-1} Q^T}} \tag{16}$$

where $\hat{\boldsymbol{\mu}}(t) = (\boldsymbol{Tr}(\lfloor nt \rfloor)^T \boldsymbol{Tr}(\lfloor nt \rfloor))^{-1} \boldsymbol{Tr}(\lfloor nt \rfloor)^T \mathbf{Y}(\lfloor nt \rfloor)$,
$\hat{\sigma}(t)^2 =$

$$[\boldsymbol{Y}(\lfloor nt \rfloor) - \boldsymbol{Tr}(\lfloor nt \rfloor)\hat{\boldsymbol{\mu}}(t)]^T [\boldsymbol{Y}(\lfloor nt \rfloor) - \boldsymbol{Tr}(\lfloor nt \rfloor)\hat{\boldsymbol{\mu}}(t)]/(\lfloor nt \rfloor - 2).$$

Note that Z_t^{tt} is just the t-test statistics.

In Table 2, we found that the type I error rates are all conservative for SPB and PS when both covariates are either discrete or continuous. In addition, when both covariates are continuous, the type I error rates are more conservative. This phenomenon depends on the parameter settings for the generated discrete or continuous covariates. However, it is worth noting that both $\sigma_{\delta j}$ and δ_j in (5) are 0 if all the covariates are discrete.

Next, we evaluate our proposed sequential monitoring and SSR methods with the sequential test statistics U_t^{ttest} defined in (11) and report the results in Table 3. All the other settings are the same as in Table 1. To implement our method, a variety of approaches such as bootstraps can be used to obtain $\hat{\epsilon}$. In this paper, we obtain $\hat{\epsilon}$ as follows. At each interim look, we fit model (14) with full data to obtain consistent estimators of the unknown parameters. We can also easily obtain consistent estimators of $\sigma_{\delta j}$ and $Var(V_j)$ based on the observed covariates due to the law of large numbers. Thus the consistency of $\hat{\epsilon}$ follows the fundamental large-sample theory [16].

From Table 3, we can see that our methods can control the type I error rate well. Same as in Table 1, SPB can lead to significantly better overall balance and stratum-level balance than complete randomization, and it also outperforms the Pocock and Simon's design. Compared to complete randomization, Pocock and Simon's design also leads to better overall, stratum-level, and marginal balance.

In Table 4, we study our methods under H_1. We set $\beta_0 = \beta_1 = \beta_2 = 1$ and vary the other parameters. Other settings are the same as in Table 3. We can see that our proposed methods can greatly increase the power, even when the power for complete randomization is already as high as 0.870. We also found that SPB can lead to better overall and stratum-level balance than CR and SPB can outperform the Pocock and Simon's design in terms of these two level balances. Besides, the Pocock and Simon's design returns a much-improved marginal balance than complete randomization.

5. Conclusion

CAR such as the stratified permuted block randomization [28] and Pocock and Simon's design [20] is one of the most popular randomization designs in Phase III confirmatory clinical trials. Due to ethical, administrative, and economic reasons, sequential monitoring is desirable in such large clinical trials. Sample size re-estimation is often necessary to guarantee the power of the trial. However, there is no comprehensive theoretical study on sequential monitoring of covariate-adaptive clinical trials with sample size re-estimation because all the three procedures have adaptive properties and simple statistical theory based on independently and identically distributed responses is not applicable here. In this paper, we studied the theoretical and numerical properties for this complex procedure. The proposed methods can successfully control the type I error rate demonstrated by the numerical study and supported by the theoretical results. Our methods can also lead to much improved power and treatment balance across subgroups compared to traditional designs.

This paper opens the door to future research topics. First, we used the idea of Cui et al. [3] for SSR, and other SSR approaches such as Gould [7], Gould and Shih [8,9], Shih [24], and Herson and Wittes [10] can be investigated. Second, it is worth proposing other approaches to adjust the test statistics to control the type I error rate. Bootstrap is a natural idea to study since the conservativeness of the type I error rate comes from a wrong estimation of the variance of the estimator of treatment difference [23]. Other methods leading to an unbiased estimate of this variance can also be investigated. Third, we focus on linear regression models in this paper. Yin and Shen [27] studied sample size calculation for correlated observations that often arise in biomedical research. Varadhan and Wang [25] proposed methods and models to investigate whether the treatment is similarly efficacious in subgroups. Besides, a large number of confirmatory clinical trials are longitudinal trials. Studying these models and scenarios for our procedure will be desirable. Forth, there are different types of adaptive randomization designs. For example, Zhang et al. [29] proposed the covariate-adjusted response-adaptive randomization (CARA) that takes into account all the previous treatment assignments, responses, covariates, and the current covariate to achieve different ethical and efficient aims. Du et al. [5] compared three response adaptive randomization designs that adaptively change the probability of treatment assignments according to how well the treatments are performing in the trial. The study on sequential monitoring of clinical trials with these adaptive randomization designs and SSR is lacking in the literature. We leave all these for future research topics.

Table 3
Performance of different designs under H_0 when adjusted t-test is used.

Both covariates are discrete										
(β_0, p_1, p_2)	Design	α	DIF_n	DIF_{11}	DIF_{10}	DIF_{01}	DIF_{00}	$DIF_{1\cdot}$	$DIF_{0\cdot}$	$DIF_{\cdot 1}$	$DIF_{\cdot 0}$
(0.5,0.5,0.5)	SPB	0.052	1.32 (1.27)	0.67 (0.62)	0.67 (0.63)	0.68 (0.62)	0.67 (0.62)				
(0.5,0.5,0.5)	PS	0.051	1.68 (1.69)	5.43 (4.16)	5.43 (4.17)	5.39 (4.17)	5.41 (4.18)	1.51 (1.37)	1.55 (1.41)	1.53 (1.41)	1.53 (1.39)
(0.5,0.5,0.5)	CR	0.049	20.8 (15.8)	10.3 (7.90)	10.5 (7.97)	10.6 (8.08)	10.5 (8.06)	14.6 (11.1)	14.9 (11.4)	14.8 (11.3)	14.8 (11.3)
(0.5,0.4,0.6)	SPB	0.053	1.31 (1.28)	0.66 (0.62)	0.67 (0.63)	0.67 (0.62)	0.66 (0.63)				
(0.5,0.4,0.6)	PS	0.054	1.67 (1.65)	5.18 (3.97)	5.17 (3.96)	5.23 (4.01)	5.22 (4.02)	1.52 (1.38)	1.52 (1.37)	1.53 (1.38)	1.52 (1.36)
(0.5,0.4,0.6)	CR	0.048	20.9 (16.1)	10.3 (7.84)	8.35 (6.46)	12.4 (9.40)	10.2 (7.85)	13.3 (10.2)	16.2 (12.3)	16.1 (12.4)	13.1 (10.2)
(2,0.5,0.5)	SPB	0.051	1.32 (1.28)	0.67 (0.63)	0.66 (0.62)	0.67 (0.62)	0.66 (0.62)				
(2,0.5,0.5)	PS	0.054	1.67 (1.67)	5.41 (4.11)	5.41 (4.11)	5.42 (4.13)	5.40 (4.09)	1.52 (1.40)	1.53 (1.37)	1.51 (1.40)	1.54 (1.41)
(2,0.5,0.5)	CR	0.050	20.9 (16.0)	10.4 (7.94)	10.6 (8.08)	10.4 (7.89)	10.4 (8.00)	14.9 (11.3)	14.7 (11.3)	14.7 (11.3)	14.9 (11.4)
(2,0.4,0.6)	SPB	0.049	1.33 (1.27)	0.66 (0.62)	0.68 (0.63)	0.67 (0.63)	0.66 (0.62)				
(2,0.4,0.6)	PS	0.055	1.68 (1.68)	5.21 (3.98)	5.16 (3.93)	5.24 (3.99)	5.22 (3.97)	1.52 (1.38)	1.52 (1.37)	1.53 (1.37)	1.52 (1.39)
(2,0.4,0.6)	CR	0.051	20.9 (16.1)	10.3 (7.83)	8.42 (6.39)	12.5 (9.53)	10.3 (7.81)	13.3 (10.1)	16.3 (12.3)	16.2 (12.4)	13.3 (10.2)
Both covariates are continuous											
(β_0, q_1, q_2)	Design	α	DIF_n	DIF_{11}	DIF_{10}	DIF_{01}	DIF_{00}	$DIF_{1\cdot}$	$DIF_{0\cdot}$	$DIF_{\cdot 1}$	$DIF_{\cdot 0}$
(0.5,0.5,0.5)	SPB	0.050	1.33 (1.27)	0.68 (0.62)	0.67 (0.62)	0.66 (0.62)	0.67 (0.63)				
(0.5,0.5,0.5)	PS	0.055	1.70 (1.67)	5.45 (4.18)	5.45 (4.17)	5.45 (4.16)	5.45 (4.16)	1.54 (1.39)	1.54 (1.39)	1.53 (1.41)	1.54 (1.38)
(0.5,0.5,0.5)	CR	0.051	20.8 (16.0)	10.5 (8.01)	10.5 (8.02)	10.4 (7.98)	10.4 (8.02)	14.8 (11.4)	14.7 (11.3)	14.8 (11.3)	14.8 (11.4)
(0.5,0.4,0.6)	SPB	0.051	1.32 (1.27)	0.68 (0.63)	0.67 (0.63)	0.67 (0.62	0.67 (0.63)				
(0.5,0.4,0.6)	PS	0.060	1.69 (1.67)	5.20 (4.02)	5.16 (3.99)	5.23 (4.05)	5.19 (4.01)	1.54 (1.38)	1.57 (1.41)	1.54 (1.41)	1.53 (1.37)
(0.5,0.4,0.6)	CR	0.047	20.9 (16.2)	10.3 (7.87)	8.40 (6.39)	12.5 (9.67)	10.3 (7.77)	13.3 (10.2)	16.2 (12.4)	16.3 (12.4)	13.3 (10.1)
(2,0.5,0.5)	SPB	0.050	1.31 (1.26)	0.67 (0.63)	0.66 (0.62)	0.66 (0.62)	0.67 (0.63)				
(2,0.5,0.5)	PS	0.054	1.68 (1.66)	5.43 (4.20)	5.42 (4.15)	5.41 (4.18)	5.39 (4.16)	1.53 (1.39)	1.52 (1.39)	1.50 (1.36)	1.53 (1.37)
(2,0.5,0.5)	CR	0.052	20.9 (16.0)	10.3 (7.91)	10.4 (7.99)	10.5 (8.04)	10.5 (7.90)	14.8 (11.4)	14.8 (11.2)	14.8 (11.3)	14.8 (11.2)
(2,0.4,0.6)	SPB	0.054	1.32 (1.28)	0.66 (0.63)	0.67 (0.63)	0.67 (0.62)	0.67 (0.62)				
(2,0.4,0.6)	PS	0.056	1.70 (1.68)	5.20 (3.96)	5.14 (3.92)	5.21 (3.98)	5.17 (3.95)	1.53 (1.42)	1.51 (1.38)	1.53 (1.38)	1.55 (1.40)
(2,0.4,0.6)	CR	0.048	21.2 (16.1)	10.0 (7.77)	8.41 (6.36)	12.6 (9.55)	10.3 (7.88)	13.2 (10.0)	16.4 (12.6)	16.2 (12.2)	13.3 (10.2)

Note: α: type I error rate; DIF_n: overall difference in patient numbers between the two treatments; DIF_{gh} for $X_1 = g$ and $X_2 = h$, g, $h = 0$, 1: the differences of patient numbers between the two treatments in the four stratum; $DIF_{1\cdot}$: marginal imbalance for $X_1 = 1$; $DIF_{0\cdot}$: marginal imbalance for $X_1 = 0$; $DIF_{\cdot 1}$: marginal imbalance for $X_2 = 1$; $DIF_{\cdot 0}$: marginal imbalance for $X_2 = 0$.

Table 4
Performance of different designs under H_1 when adjusted t-test is used.

Both covariates are discrete											
(β_T, p_1, p_2)	Design	Power	DIF_n	DIF_{11}	DIF_{10}	DIF_{01}	DIF_{00}	$DIF_{1\cdot}$	$DIF_{0\cdot}$	$DIF_{\cdot 1}$	$DIF_{\cdot 0}$
(0.3,0.5,0.5)	SPB	0.893	1.33 (1.26)	0.65 (0.62)	0.68 (0.63)	0.67 (0.62)	0.67 (0.62)				
(0.3,0.5,0.5)	PS	0.892	1.69 (1.67)	4.46 (3.54)	4.45 (3.55)	4.44 (3.55)	4.46 (3.57)	1.53 (1.40)	1.54 (1.39)	1.51 (1.35)	1.53 (1.40)
(0.3,0.5,0.5)	CR	0.735	18.2 (14.3)	9.21 (7.16)	9.21 (7.19)	9.24 (7.32)	9.21 (7.15)	12.9 (10.1)	13.0 (10.2)	13.0 (10.3)	13.0 (10.2)
(0.3,0.4,0.6)	SPB	0.892	1.30 (1.25)	0.66 (0.62)	0.67 (0.62)	0.66 (0.62)	0.67 (0.63)				
(0.3,0.4,0.6)	PS	0.891	1.65 (1.64)	4.38 (3.44)	4.32 (3.42)	4.41 (3.45)	4.37 (3.46)	1.53 (1.38)	1.49 (1.34)	1.54 (1.38)	1.53 (1.36)
(0.3,0.4,0.6)	CR	0.747	18.2 (14.4)	8.93 (6.98)	7.28 (5.78)	11.0 (8.65)	9.03 (7.05)	11.6 (9.12)	14.3 (11.1)	14.1 (11.0)	11.7 (9.07)
(0.35,0.5,0.5)	SPB	0.961	1.32 (1.28)	0.67 (0.62)	0.66 (0.62)	0.67 (0.63)	0.68 (0.63)				
(0.35,0.5,0.5)	PS	0.959	1.66 (1.66)	4.21 (3.33)	4.21 (3.31)	4.20 (3.34)	4.16 (3.30)	1.53 (1.37)	1.52 (1.37)	1.51 (1.36)	1.52 (1.38)
(0.35,0.5,0.5)	CR	0.860	17.4 (13.9)	8.93 (6.94)	8.75 (6.90)	8.80 (7.05)	8.72 (6.87)	12.4 (9.75)	12.4 (9.72)	12.6 (9.85)	12.2 (9.71)
(0.35,0.4,0.6)	SPB	0.960	1.30 (1.25)	0.67 (0.63)	0.67 (0.62)	0.67 (0.63)	0.65 (0.63)				
(0.35,0.4,0.6)	PS	0.965	1.71 (1.67)	4.01 (3.20)	3.93 (3.15)	4.06 (3.23)	4.00 (3.19)	1.54 (1.39)	1.53 (1.38)	1.53 (1.38)	1.53 (1.40)
(0.35,0.4,0.6)	CR	0.870	17.4 (13.7)	8.48 (6.71)	7.02 (5.59)	10.3 (8.13)	8.61 (6.71)	11.0 (8.66)	13.4 (10.4)	13.3 (10.6)	11.2 (8.76)
Both covariates are continuous											
(β_T, q_1, q_2)	Design	Power	DIF_n	DIF_{11}	DIF_{10}	DIF_{01}	DIF_{00}	$DIF_{1\cdot}$	$DIF_{0\cdot}$	$DIF_{\cdot 1}$	$DIF_{\cdot 0}$
(0.4,0.5,0.5)	SPB	0.903	1.32 (1.27)	0.67 (0.62)	0.67 (0.63)	0.67 (0.63)	0.67 (0.63)				
(0.4,0.5,0.5)	PS	0.900	1.71 (1.69)	4.46 (3.55)	4.47 (3.57)	4.44 (3.56)	4.45 (3.55)	1.55 (1.41)	1.51 (1.38)	1.55 (1.41)	1.53 (1.38)
(0.4,0.5,0.5)	CR	0.691	18.6 (14.6)	9.36 (7.22)	9.34 (7.26)	9.39 (7.28)	9.38 (7.29)	13.3 (10.3)	13.3 (10.3)	13.2 (10.2)	13.2 (10.3)
(0.4,0.4,0.6)	SPB	0.889	1.33 (1.28)	0.67 (0.62)	0.68 (0.63)	0.66 (0.62)	0.68 (0.62)				
(0.4,0.4,0.6)	PS	0.895	1.68 (1.67)	4.33 (3.44)	4.27 (3.39)	4.39 (3.45)	4.33 (3.43)	1.51 (1.36)	1.54 (1.39)	1.53 (1.39)	1.52 (1.39)
(0.4,0.4,0.6)	CR	0.688	18.5 (14.5)	9.07 (7.05)	7.49 (5.85)	11.2 (8.66)	9.21 (7.16)	11.8 (9.15)	14.4 (11.3)	14.4 (11.1)	11.9 (9.15)
(0.45,0.5,0.5)	SPB	0.952	1.30 (1.27)	1.30 (1.27)	0.67 (0.62)	0.67 (0.62)	0.67 (0.63)	0.66 (0.62)			
(0.45,0.5,0.5)	PS	0.952	1.68 (1.67)	4.25 (3.36)	4.23 (3.36)	4.29 (3.38)	4.25 (3.36)	1.51 (1.36)	1.52 (1.37)	1.52 (1.38)	1.51 (1.39)
(0.45,0.5,0.5)	CR	0.784	18.0 (14.1)	9.15 (6.98)	9.00 (7.17)	9.14 (7.17)	9.03 (6.97)	12.7 (9.95)	12.75 (10.0)	12.8 (10.0)	12.8 (9.96)
(0.45,0.4,0.6)	SPB	0.953	1.30 (1.26)	0.67 (0.63)	0.67 (0.62)	0.67 (0.62)	0.66 (0.62)				
(0.45,0.4,0.6)	PS	0.948	1.70 (1.67)	4.11 (3.29)	4.01 (3.26)	4.14 (3.29)	4.07 (3.27)	1.54 (1.40)	1.54 (1.41)	1.55 (1.40)	1.54 (1.39)
(0.45,0.4,0.6)	CR	0.792	18.2 (14.2)	8.82 (6.80)	7.15 (5.60)	10.9 (8.53)	8.99 (6.93)	11.3 (8.86)	14.1 (11.0)	14.1 (11.0)	11.5 (9.02)

Note: DIF_n: overall difference in patient numbers between the two treatments; DIF_{gh} for $X_1 = g$ and $X_2 = h$, g, $h = 0$, 1: the differences of patient numbers between the two treatments in the four stratum; $DIF_{1\cdot}$: marginal imbalance for $X_1 = 1$; $DIF_{0\cdot}$: marginal imbalance for $X_1 = 0$; $DIF_{\cdot 1}$: marginal imbalance for $X_2 = 1$; $DIF_{\cdot 0}$: marginal imbalance for $X_2 = 0$.

Acknowledgements

This work was partially supported by Cancer Prevention Research Institute of Texas (RP170668).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cct.2019.105874.

References

[1] A. Baldi Antognini, M. Zagoraiou, The covariate-adaptive biased coin design for balancing clinical trials in the presence of prognostic factors, Biometrika 98 (3) (2011) 519–535.
[2] P.J. Barnes, S.J. Pocock, H. Magnussen, A. Iqbal, B. Kramer, M. Higgins, D. Lawrence, Integrating indacaterol dose selection in a clinical study in COPD using an adaptive seamless design, Pulm. Pharmacol. Ther. 23 (3) (2010) 165–171.
[3] L. Cui, H.M. Hung, S.J. Wang, Modification of sample size in group sequential clinical trials, Biometrics. 55 (1999) 853–857.
[4] J.S. Denne, Sequential Procedures for Sample Size Estimation, (1996) (Doctoral dissertation, University of Bath).
[5] Y. Du, X. Wang, J.J. Lee, Simulation study for evaluating the performance of response-adaptive randomization, Contemporary Clinical Trials 40 (2015) 15–25.
[6] FDA, Draft Guidance for Industry: Adaptive Design Clinical Trials for Drugs and Biologics, (2010).
[7] A.L. Gould, Interim analyses for monitoring clinical trials that do not materially affect the type I error rate, Stat. Med. 11 (1992) 55–66.
[8] A.L. Gould, W.J. Shih, Sample size re-estimation without unblinding for normally distributed outcomes with unknown variance, Communications in Statistics-Theory and Methods. 21 (1992) 2833–2853.
[9] A.L. Gould, W.J. Shih, Modifying the design of ongoing trials without unblinding, Stat. Med. 17 (1998) 89–100.
[10] J. Herson, J. Wittes, The use of interim analysis for sample size adjustment, Drug Information Journal. 27 (1993) 753–760.
[11] Y. Hu, F. Hu, Asymptotic properties of covariate-adaptive randomization, Ann. Stat. 40 (2012) 1794–1815.
[12] A.T. Iacono, 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, A randomized trial of inhaled cyclosporine in lung-transplant recipients, N. Engl. J. Med. 354 (2) (2006) 141–150.
[13] C. Jennison, B.W. Turnbull, Group Sequential Methods with Applications to Clinical Trials, Chapman and Hall/CRC, 2000.
[14] H.L. Lai, C.J. Chen, T.C. Peng, F.M. Chang, M.L. Hsieh, H.Y. Huang, S.C. Chang, Randomized controlled trial of music during kangaroo care on maternal state anxiety and preterm infants' responses, Int. J. Nurs. Stud. 43 (2006) 139–146.
[15] K. Lan, D.L. DeMets, Discrete sequential boundaries for clinical trials, Biometrika 70 (1983) 659–663.
[16] E.L. Lehmann, Elements of Large-Sample Theory, Springer Science & Business Media, 2004.
[17] W. Ma, F. Hu, L. Zhang, Testing hypotheses of covariate-adaptive randomized clinical trials, J. Am. Stat. Assoc. 110 (2015) 669–680.
[18] P.C. O'Brien, T.R. Fleming, A multiple testing procedure for clinical trials, Biometrics 35 (1979) 549–556.
[19] S.J. Pocock, Group sequential methods in the design and analysis of clinical trials, Biometrika 64 (1977) 191–199.
[20] S. Pocock, R. Simon, Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial, Biometrics 31 (1975) 103–115.
[21] M.A. Proschan, K. Lan, J.T. Wittes, Statistical Monitoring of Clinical Trials, a Unified Approach, Springer Science + Business Media, LLC, 2006.
[22] W.F. Rosenberger, J.M. Lachin, Randomization in Clinical Trials: Theory and Practice, John Wiley and Sons, 2015.
[23] J. Shao, X. Yu, B. Zhong, A theory for testing hypotheses under covariate-adaptive randomization, Biometrika 97 (2010) 347–360.
[24] W.J. Shih, Sample size re-estimation in clinical trials, in: K.E. Peace (Ed.), Biopharmaceutical Sequential Statistical Applications, Marcel Dekker, New York, 1992, pp. 285–301.
[25] R. Varadhan, S.J. Wang, Standardization for subgroup analysis in randomized controlled trials, J. Biopharm. Stat. 24 (1) (2014) 154–167.
[26] A. Wald, Sequential Analysis, John Wiley and Sons Inc., New York, 1947.
[27] G. Yin, Y. Shen, Adaptive design and estimation in randomized clinical trials with correlated observations, Biometrics 61 (2) (2005) 362–369.
[28] M. Zelen, The randomization and stratification of patients to clinical trials, J. Chronic Dis. 27 (7) (1974) 365–375.
[29] L.X. Zhang, F. Hu, S.H. Cheung, W.S. Chan, Asymptotic properties of covariate-adjusted response-adaptive designs, Ann. Stat. 35 (3) (2007) 1166–1182.
[30] H. Zhu, F. Hu, Sequential monitoring of covariate-adaptive randomized clinical trials, Stat. Sin. 29 (2018) 265–282.