Fall 12-2018

SEQUENTIAL MONITORING OF ADAPTIVE RANDOMIZED CLINICAL TRIALS WITH SAMPLE SIZE RE-ESTIMATION

Jun Yu
UTHealth School of Public Health

SEQUENTIAL MONITORING OF ADAPTIVE RANDOMIZED CLINICAL TRIALS WITH SAMPLE SIZE RE-ESTIMATION

by

JUN YU, MD EQUIVALENT, MS

APPROVED:

_____

RUOSHA LI, PHD

_____

DEJIAN LAI, PHD

_____

HAN CHEN, PHD

_____

J. MICHAEL SWINT, PHD

_____

DEAN, THE UNIVERSITY OF TEXAS SCHOOL OF PUBLIC HEALTH

DEDICATION

To my dearest family

SEQUENTIAL MONITORING OF ADAPTIVE RANDOMIZED CLINICAL TRIALS WITH SAMPLE SIZE RE-ESTIMATION

by

JUN YU

MD equivalent, Shandong University, China, 2006
MS, University of Virginia, USA, 2008

Presented to the Faculty of The University of Texas
School of Public Health
in Partial Fulfillment
of the Requirements
for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS
SCHOOL OF PUBLIC HEALTH
Houston, Texas
December, 2018

ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincere gratitude to my dissertation advisor, Dr. Dejian Lai for his excellent guidance and valuable advice on both the statistical knowledge and computational technology. Without his persistent support, this dissertation would not have been worked out. In the mean time, I am deeply grateful to Dr. Ruosha Li for serving as my academic advisor. I thank her for constantly monitoring my academic progress during the doctorate training and providing thoughtful comments on my dissertation. I thank Dr. J. Michael Swint for being my minor advisor in Health Economics to give suggestions on course selection for fulfilling minor requirement. I thank Dr. Han Chen for being my breadth advisor and configuring course development for breadth requirement. I also thank Dr. Chunyan Cai for being my external reviewer for both the dissertation proposal defense and the final dissertation defense. A special thanks goes to Dr. Wei Ma for his valuable comments and time.

My deepest gratitude goes to my parents, Renzhong Yu and Fei Shan, my husband, Zhu Zhu, my children, Yuqi and Yulin, for their dedicated and unconditional love, encouragement and always being there for me. Thank you for supporting me through this entire journey, and helping me becoming a better person. You are my motivation to keep on fighting and never give up. I am so lucky to have your support.

SEQUENTIAL MONITORING OF ADAPTIVE RANDOMIZED CLINICAL TRIALS WITH SAMPLE SIZE RE-ESTIMATION

Jun Yu, MD equivalent, MS, PhD

The University of Texas
School of Public Health, 2018

Dissertation Chair: Dejian Lai, PhD

Clinical trials are complicated and involve human beings. Therefore, lots of ethical and efficient objectives are expected to be achieved. These objectives include maximizing the power of detecting the treatment effects, assigning more patients to the better treatments, saving the cost and time, and controlling the type I error rate. A variety of adaptive designs have been proposed to achieve different aims, among which sequential monitoring and sample size re-estimation are very popular in real clinical trials. In addition, adaptive randomization designs sequentially update the allocation probability aiming to target different allocation proportions and achieve different aims. Hu and Rosenberger (2006) classified adaptive randomization design into four categories, i.e., permuted block randomization, covariate-adaptive randomization (CAR), response-adaptive randomization (RAR), and covariate-adjusted response-adaptive randomization. In this dissertation, I investigate the combination of sequential monitoring, sample size re-estimation, and two types of adaptive randomization designs, i.e., CAR and RAR. For RAR, I focus on urn models. For CAR, I study three scenarios depending on whether all, part, or none of the randomization covariates are included in the data analysis. I propose methods to control the type I error rate, offer the theoretical results, and perform comprehensive numerical studies to show that the methods can protect the type I error rate and have advantages over traditional designs.

TABLE OF CONTENTS

LIST OF TABLES

Chapter 1

Introduction

In this dissertation, I investigated the implementation of response-adaptive randomization (RAR) procedures in real clinical trials by investigating the sequential monitoring of response-adaptive randomized clinical trials with sample size re-estimation (SSR) (Chapter 2). I also theoretically and numerically studied the combination of sequential monitoring, SSR, and covariate-adaptive randomization (Chapters 3 and 4). Such combination is the most popular procedure in Phase III confirmatory clinical trials, but its theoretical investigation is lacking in the literature. In this chapter, I introduce my research in the context of clinical trials.

1.1 Significance

Randomized controlled clinical trials (RCTs) are the gold standard for evaluating the efficacy and safety of new drugs for approval (Friedman et al., 2015). RCTs are very complicated system involving planning, conducting, analyzing and assessing. Conventionally a two-arm (a new drug arm versus a control arm) trial is well planned. An assumed effect size, a level of significance and a study power are essential to determine the sample size of clinical trials. After the sample size estimation, the patients enrollment

starts, followed by a particular randomization procedure to allocate the two arms. A conclusion of whether the trial is positive or not is made based on the carefully collected and analyzed data at the end of the trial. This traditional procedure has several problems and corresponding solutions have been proposed. I give a brief review as follows.

1.1.1 Covariate-adaptive randomization

In clinical trials, covariates often play an important role. Some covariates are known to be important risk factors associated with the response of a patient to the treatment. In a randomized trial, it is crucial to balance such covariates in each of the treatment arms so as to avoid the biases introduced into the estimation of treatment effect due to covariate imbalance. Covariate-adaptive randomization (CAR) designs have been proposed to achieve this aim (Rosenberger and Lachin, 2015).

CAR sequentially allocates patients based on previous treatment assignments and covariates, and the covariates of the current patient. The most commonly used CAR designs include stratified permuted block randomization design (Zelen, 1974), Pocock and Simon's procedure (1975). However, there are two problems. Firstly, the treatment assignments and responses from CAR are not independently and identically distributed any more due to the randomization mechanism (Hu and Hu, 2012; Ma et al., 2015). Secondly, usually in practice, not all covariates used in the randomization can be completely utilized in the inference procedures. For instance, in a clinical trial described in Anderson et al. (2000), the Pocock and Simon's procedure was applied to balance allocation over three covariates including disease extent, performance status, and clinical centers. Nevertheless, a two-sample t-test was conducted to compare a continuous primary endpoint between two treatment groups, without adjusting any covariate effects. The reasons why some covariates used in randomization are neglected in final analysis include: (i) controlling for too many covariates means complicated modeling methodology; (ii) it is

hard to interpret some covariates in the analysis model (e.g., clinical centers, etc); (iii) the justification of the model specification becomes more difficult if more covariates are included in the model. Concerns are raised about the validity of statistical inference for CAR designs. Birkett (1985) and Forsythe et al. (1987) found that the two-sample t-test is conservative in terms of Type I error if Taves' minimization is utilized to allocate patients to treatments through simulation studies. In practice, conventional tests are often conducted without consideration of CAR scheme. 'Conservative' means that the observed type I error rate is smaller than the nominal type I error rate. It remains a concern if conventional tests are still valid under CAR designs.

1.1.2 Response-adaptive randomization

Balance is not always the optimal allocation proportions in terms of certain objectives. For example, when comparing the mean of two normal distributions, $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$, the following Neyman allocation (Neyman, 1934) instead of the equal allocation is the optimal one in terms of power when the variances of the two distribution are not the same,

$$\rho_1 = \frac{\sigma_1}{\sigma_1 + \sigma_2}, \ \rho_2 = 1 - \rho_1, \tag{1.1}$$

where ρ_1 and ρ_2 are the allocation proportion to treatments 1 and 2, respectively. For binary responses, the famous optimal allocation proportions with corresponding objectives are listed below

(1) **(Neyman allocation)** Objective: Maximizing the power.

$$\rho_1 = \frac{\sqrt{p_1(1-p_1)}}{\sqrt{p_1(1-p_1)} + \sqrt{p_2(1-p_2)}}, \ \rho_2 = 1 - \rho_1. \tag{1.2}$$

(2) **(Optimal allocation)** Objective: Minimizing the expected number of failures

while fixing power. (Rosenberger et al., 2001)

$$\rho_1 = \frac{\sqrt{p_1}}{\sqrt{p_1} + \sqrt{p_2}}, \ \rho_2 = 1 - \rho_1. \tag{1.3}$$

(3) **(Urn Allocation)** Objective: Assigning more patients to the better treatment.

$$\rho_1 = \frac{1 - p_2}{(1 - p_1) + (1 - p_2)}, \ \rho_2 = 1 - \rho_1, \tag{1.4}$$

where p_1 and p_2 as the success rates for two treatments, respectively. Response-adaptive randomization procedures sequentially update the allocation probability of the next patient based on all the previous treatment assignments and responses in order to achieve ethical and efficient objectives such as maximizing the power to detect the treatment difference, minimizing the total numbers of failures, etc. There are three steps to implement the optimal RAR procedures in practice. First, we determine the main objectives and mathematically formulate these objectives. Second, we derive the target allocation proportion which achieves these objectives. Third, we implement certain RAR procedures to target the optimal allocation proportion.

1.1.3 Sequential monitoring

In clinical trials, it is not rare to perform interim analyses based on accrued data up to a certain time point during the conduct of a clinical trial due to ethical consideration, administrative reasons, and economic constraints (Jennison and Turnbull, 2000). Sequential methods usually lead to savings in sample size, cost and time when compared with the other fixed sample designs. A group sequential test is referred to as a test performed based on accrued data at some pre-specified intervals rather than after every new observation is obtained (Jennison and Turnbull, 2000). For a sequential trial

with multiple interim analyses, multiple tests cause an inflation of the type I error rate, so it is necessary to adjust α-level at each interim analysis. Other research on group sequential designs can be seen in Simon (1989), Ensign et al. (1994), Chen (1997), Chen and Ng (1998), Sargent and Goldberg (2001), Wu and Lan (1992), Lan and DeMets (1983), Wang and Tsiatis (1987), Proschan et al. (2006), Pocock (1977), and O'Brien and Fleming (1979).

1.1.4 Sample size re-estimation

In clinical trials, a sufficient number of sample size is necessary to reach a desired power for detecting a treatment difference of clinical importance, if such a meaningful difference truly exists. To achieve this aim, the number of the required subjects is estimated under certain assumptions by a power analysis at the planning stage of the trial. The sample size estimation of the pre-study power analysis is usually based on the assumed treatment effect. However, the true treatment effect may be different from the initial assumption, therefore the study is possibly over-powered or under-powered. Thus, to re-estimate sample sizes adaptively based on observed data in an interim analysis is of interest (Chow et al., 2008; Lehmacher and Wassmer, 1999; Cui et al., 1999; Mehta and Pocock, 2011; Lai, 2013).

1.2 Literature review

This dissertation studies the combination of three types of adaptive designs. I conduct a brief literature review starting from the general concept, adaptive design.

Adaptive designs utilize accumulating data to adjust the clinical trial procedures without undermining the validity and integrity of the trials. The validity includes internal and external validity. Internal validity is the reasonable representation of the

treatment effects within the study population. Basically, if the treatment differences are detected, we will ask whether the differences are due to the treatments, patient characteristics, or chance. If no treatment effects are detected, we would like to ask whether it is due to the true equivalence, misconduct, or lack of precision (study power). To support the internal validity, we need to design trials including comparable groups, and try to avoid or minimize biases in the treatment allocation, assessing treatment effects, study monitoring and data analysis, and multiple hypothesis testing. These biases can be minimized by appropriate randomization and stratification, using concurrent control group and masking the treatment assignment, performing ongoing review by disciplined investigators and expert statisticians, and predefining hypotheses and endpoints in the protocol (Shih and Aisner, 2015). External validity is the validity of inferences as they pertain to the generalizability to future subjects (Rothwell, 2005). In the study protocol, the patient characteristics, treatment and procedures, outcome measures, and follow-up together define the generalizability and applicability of the trial results. For supporting external validity, the later phases of a clinical trial should be conducted by multiple investigators in different medical settings, including university teaching hospitals, community medical centers, private clinics, etc., as well as in various geographical regions.

Chang (2014) classified adaptive designs into the following categories: (1) *group sequential designs* (2) *sample size re-estimation* (3) *drop-losers designs* (4) *adaptive randomization design* (5) *adaptive dose-escalation designs* (6) *biomarker-adaptive designs* (7) *adaptive treatment-switching designs* (8) *combined adaptive designs.*

My dissertation studies three types of adaptive designs listed above: group sequential designs, sample size re-estimation, and adaptive randomization designs. The idea of adaptive randomization can be traced back to Thompson (1933) and Robbins (1952). Hu and Rosenberger (2006) classified adaptive randomization design into four categories, i.e., permuted block randomization, covariate-adaptive randomization (CAR),

response-adaptive randomization (RAR), and covariate-adjusted response-adaptive randomization. I study RAR in Chapter 2 and CAR in Chapters 3 and 4.

1.2.1 Covariate-adaptive randomization

To equalize the distribution of covariates within each treatment group and minimize the imbalance, many covariate-adaptive designs were proposed in the literature.

One idea is to stratify the patients according to covariates before randomization and then to employ separate randomization for each stratum. For a small set of known discrete covariates, one of the most commonly used methods is the stratified permuted block randomization design which determines the strata first with the covariates' levels and then perform the permuted block randomization within each stratum. One serious drawback of this method is that the number of strata increases quickly as the number of covariates and the number of the covariate levels increase. If the sample size is relatively small compared to the number of strata, it is almost equivalent to complete randomization, losing its advantages (Rosenberger and Lachin, 2015).

To ensure balance over a large number of covariates, there are various methods proposed to determine the treatment assignment of a new subject to minimize the covariate imbalance within each treatment group. The first covariate-adaptive design was proposed in the mid-1970s by Taves (1974). He proposed the method to minimize imbalance on key covariates. Pocock and Simon (1975) proposed generalizations of minimization to randomized clinical trials. Because they balance covariates marginally, these methods are referred as marginal procedures. For notation purposes, if discrete covariate $Z_i, i = 1,\ldots, I$ has n_i levels, then they balance on covariates within each of $\sum_{i=1}^{I} n_i$ levels of given covariates.

In the covariate-adaptive randomization procedure proposed by Pocock and Simon (1975), let $N_{ijk}(n)$, i=1,..., I, j=1,..., n_i, k=1, 2 (1=A, 2=B), be the number of patients

in stratum j of covariate i on treatment k after n patients have been randomized. Suppose the $(n+1)$th patient to be randomized is a member of strata $r_1,\ldots,r_I$ of covariates 1,..., I. Then $D_i(n) = N_{ir_i1}(n) - N_{ir_i2}(n)$ is computed for each i=1,..., I. A weighted sum is then taken as $D(n) = \sum_{i=1}^{I}(w_i D_i(n))$, where w_i are weights chosen depending on which covariates are deemed of greater importance. The measure $D(n)$ is used to determine the allocation probability of the $(n+1)$th patient. If $D(n) > 0$ (< 0), then one decreases (increases) the probability of being assigned to treatment 1 accordingly. Pocock and Simon (1975) formulated a general rule using Efron's (1971) biased coin design as

$$\phi_{n+1} = \begin{cases} 1/2, & \text{if} \quad D(n) = 0, \\ p, & \text{if} \quad D(n) < 0, \\ 1-p, & \text{if} \quad D(n) > 0. \end{cases}$$

When p = 1, we have Taves's (1974) minimization method, which is non-randomized. Pocock and Simon (1975) investigated p = 3/4.

Both stratified permuted block design and Pocock and Simon's marginal method are widely implemented in clinical research. Stratified permuted block design was employed in many clinical trials, including Iacono et al. (2006) and Jakob et al. (2012). According to Taves (2010), Pocock and Simon's marginal procedure was applied in over 400 clinical trials from 1989 to 2008. Some recent examples include Anderson et al. (2000), Gridelli et al. (2003), Krueger et al. (2007), Molander et al. (2007), Ohtori et al. (2012), etc. Hu and Hu (2012) raised some limitations of these traditional approaches and proposed a generalized family of covariate-adaptive designs along with their theoretical properties. For more discussion of handling covariates in clinical trials, see McEntegart (2003), Rosenberger and Sverdlov (2008).

Nowadays, it is widely accepted that all covariates utilized in the CAR design should be incorporated in statistical inference procedures (Ma et al., 2015). Feinstein and

Landis (1976) and Green and Byar (1978) explored the inference problems for stratified randomization for binary responses. Forsythe et al. (1987) suggested that all covariates utilized in minimization method should be included into analysis. Shao et al. (2010) theoretically proved that, the two-sample t-test is conservative under the covariate-adaptive biased coin procedure, by assuming that the response primarily follows a simple homogeneous linear model. More discussions can be found in Tu et al. (2000), Aickin (2009), and so on.

1.2.2 Response-adaptive randomization

Zelen (1969) proposed the play-the-winner rule for comparing two treatments with binary responses in clinical trials. If the response of the current patient is a success, then the same treatment will be given to next patient. If the response of the current patient is a failure, then the other treatment will be given to the next patient. With play-the-winner rule, more patients will be assigned to the better treatment. But it is a deterministic design, and a variety of bias could be introduced. The idea of incorporating randomization in the context of RAR designs stemmed from the randomized-play-the-winner rule proposed by Wei and Durham (1978). In general, there are two main families of RAR procedures: doubly-adaptive biased coin designs that is based on certain optimal criteria and urn models based on intuitive motivation. Next I will introduce the DBCD and urn-model based randomization procedures respectively.

Doubly-adaptive biased coin design

We start from the Efron's biased coin design for balancing the experiment and mitigate various forms of bias at the same time. Let $N_j(i), i = 1, 2, \ldots, j = 1, 2$ be the number of patients assigned to treatment j after the ith patient have been enrolled and assigned to treatments. The Efron's procedure sequentially assigns the next patient to treatment 1

with probability

$$\begin{aligned}\phi_{i1} &= 1/2 \text{ if } D_{i-1} = 0,\\ &= \pi \text{ if } D_{i-1} < 0,\\ &= 1-\pi \text{ if } D_{i-1} > 0,\end{aligned}$$

where $D_i = N_1(i) - N_2(i)$ is the imbalance between treatment 1 and 2 and $\pi \in (0.5, 1]$.

Balance is not always the target. Eisele (1994) and Eisele and Woodroofe (1995) proposed the doubly-adaptive biased coin design (DBCD) that sequentially assigns the next patient using both the current allocation proportions and the currently estimated optimal allocation proportion. But their conditions are very restrictive. Hu and Zhang (2004) proposed a family of DBCD and derived the asymptotic properties under widely satisfied conditions. They obtained the strong consistency, a law of the iterated logarithm and asymptotic normality of the parameter estimators. However, the procedure proposed by Hu and Zhang (2004) did not reach the asymptotic lower bound on the variability of response-adaptive designs (Hu et al., 2006). Hu et al. (2009) proposed a new family of efficient randomized adaptive designs (ERADE) that achieved the asymptotic lower bound. In this dissertation, I mainly focus on urn models below, since urn models have been used in real clinical trials (Rout et al., 1993; Bartlett et al., 1985; Tamura et al., 1994).

Urn-model based randomization

The urn models are originally in the field of probability. The Pólya urn models was proposed by Eggenberger and Pólya (1923). The initial urn contains Y_{01} balls of type 1 and Y_{02} balls of type 2. At every stage, a ball is randomly drawn and replaced and α balls of same type are added back to the urn. Friedman (1949) modified Pólya urn models by allowing adding additional β balls of the opposite type selected. Athreya and Karlin (1967, 1968) and Athreya (1969) proposed the Generalized Friedman's Urn

(GFU) as follows. When comparing K treatments, the initial urn contains balls of K types with composition $\mathbf{Y}_0 = (Y_{0,1}, \ldots, Y_{0,K})$. At stage i, $i = 1, 2, \ldots$, a ball, say type k, is drawn and replaced. Then the treatment k is assigned to the ith patient, and additional $d_i(k, g, \xi_{i,k})$ balls of type g, $g = 1, 2, \ldots, K$, are added to the urn, where $d_i(k, g, \xi_{i,k})$ is a function of $\xi_{i,k}$, the response of the ith patient assigned to treatment k. After n patients have been assigned, the urn composition is $\mathbf{Y}_n = (Y_{n,1}, \ldots, Y_{n,K})$. Define $\boldsymbol{D}_n = (d_n(k, g, \xi_{n,k}), k, g = 1, \ldots, K)$, $\boldsymbol{\xi}_n = (\xi_{n,1}, \ldots, \xi_{n,K})$ and the observed result of the nth draw $\boldsymbol{X}_n = (X_{n,1}, \ldots, X_{n,K})$. Let $\boldsymbol{H}_i = (E[d_i(k, g, \xi_{i,k})|\mathcal{F}_{i-1}], k, g = 1, \ldots, K)$, where the $\sigma-$field $\mathcal{F}_i$ is generated by $\{\boldsymbol{Y}_0, \boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_i, \boldsymbol{X}_1, \ldots, \boldsymbol{X}_i, \boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_i\}$. $\boldsymbol{D}_i$ and $\boldsymbol{H}_i$ are called the addition rules and the generating matrices, respectively.

Other important urn models in the literature include Wei (1979), Durham and Yu (1990), Smythe (1996), Durham et al. (1998), Ivanova and Rosenberger (2000), Ivanova and Flournoy (2001), Ivanova (2003), Andersen et al. (1994), Bai et al. (2002). Finally, Zhang et al. (2006) proposed the sequential estimation-adjusted urn model (SEU) and their model can target any allocation proportion and include the randomized play-the-winner rule as a special case. In this dissertation, I focus on SEU model for trials with two treatment groups.

1.2.3 Sequential monitoring

There are three primary reasons for conducting interim analysis (Jennison and Turnbull, 2000): (i) ethical consideration, (ii) administrative reasons, and (iii) economic constraints. In practice, human subjects are involved in clinical trials, so from an ethical point of view, interim analysis to make sure that the human subjects are not exposed to unnecessary negative treatments. The ineffective or unsafe trials should be terminated as early as possible to protect the subjects. From an administrative point of view, it is necessary for monitoring the trials to make sure that the clinical trials are being imple-

mented as planned. If the critical assumptions are violated, modifications or adjustment should be made so as to guarantee the integrity and quality of the trials. If the violation of the protocol is found to be much enough to fundamentally alternate the results, the trial should be stopped early. Often, clinical trials are very expensive and time consuming, so the sponsors would like to know whether there is enough ethical and statistical evidence to make the decision of stopping or continuing the trials from the economic point of view. Interim analysis usually lead to savings in sample size, cost and time when compared with the other fixed sample designs.

In the literature, there are many sequential monitoring design methods proposed, to stop the trials as early as possible when the test regimen is ineffective or unsafe, and at the same time, to avoid terminating a trial too early when the test regimen is promising. For a sequential trial with K interim analyses, the main concern is the inflation of the type I error rate, since we have more chances to reject the null hypothesis when it is true. The natural approach is to find the joint distribution of the sequential statistics, and to find corresponding critical values to control the type I error rate.

Proschan et al. (2006) introduced a unified approach for group sequential trial design. The unified approach is briefly described below. Consider a group sequential study consisting of up to K analyses. Thus, we have a sequence of test statistics $\{Z_1, \ldots, Z_K\}$. Assuming that these test statistics follow a joint canonical distribution with information levels $\{I_1, \ldots, I_k\}$ for the treatment effect. Thus, we have

$$Z_k \sim N(\theta\sqrt{I_k}, 1), k = 1, \ldots, K,$$

and

$$Cov(Z_{k1}, Z_{k2}) = \sqrt{I_{k1}/I_{k2}}, 1 \leq k_1 \leq k_2 \leq K.$$

Table 1.1 summarizes unified formulation for different types of study endpoints

under a group sequential design.

In the following, I review the general framework to determine the boundaries for early stopping of a given trial due to (i) efficacy, (ii) futility, and (iii) efficacy or futility assuming that there are a total of K analyses in the trial (Chow and Chang, 2011).

For the case of early stopping, we consider testing the one-sided null hypothesis that $H_0 : \mu_A \leq \mu_B$, where μ_A and μ_B could be means, proportions or hazard rates for treatment groups A and B, respectively.

The decision rules for early stopping for efficacy are

$$\begin{cases} \text{If } Z_k < \alpha_k, & \text{continue on next stage;} \\ \text{If } Z_k \geq \alpha_k, & \text{stop and reject } H_0, \text{ k=1,...K-1,} \end{cases}$$

and

$$\begin{cases} \text{If } Z_K < \alpha_K, & \text{stop and accept } H_0; \\ \text{If } Z_K \geq \alpha_K, & \text{stop and reject } H_0. \end{cases}$$

Wang and Tsiatis' boundary function is given by

$$\alpha_k = \alpha_K (\frac{k}{K})^{\Delta - 1/2}$$

The decision rules for early stopping for futility are

$$\begin{cases} \text{If } Z_k < \beta_k, & \text{stop and accept } H_0; \\ \text{If } Z_k \geq \beta_k, & \text{continue on next stage, k=1,...K-1,} \end{cases}$$

and

$$\begin{cases} \text{If } Z_K < \beta_K, & \text{stop and accept } H_0; \\ \text{If } Z_K \geq \beta_K, & \text{stop and reject } H_0. \end{cases}$$

The boundary function is

$$\beta_k = 2\beta_K\sqrt{\frac{k}{K}} - \beta_K(\frac{k}{K})^{\Delta-1/2}$$

The decision rules for early stopping for efficacy or futility are

$$\begin{cases} \text{If } Z_k < \beta_k, (k = 1,\ldots K), & \text{stop and accept } H_0; \\ \text{If } Z_k \geq \alpha_k, (k = 1,\ldots K), & \text{stop and reject } H_0. \end{cases}$$

The stopping boundaries are the combination of the previous efficacy and futility stopping boundaries, which is given by

$$\begin{cases} \alpha_k = \alpha_K(\frac{k}{K})^{\Delta-1/2} \\ \beta_k = 2\beta_K\sqrt{\frac{k}{K}} - \beta_K(\frac{k}{K})^{\Delta-1/2} \end{cases}$$

Lan and DeMets (1983) proposed the spending function methods to distribute (or spend) the total type I error rate as a continuous function of the information time in group sequential trial designs for interim analysis. This continuous function of the information time is referred to as the alpha spending function, denoted by $\alpha(s)$. Let s_1 and s_2 be two information times, $0 < s_1 < s_2 < 1$. Then $0 < \alpha(s_1) < \alpha(s_2) < \alpha$. $\alpha(s_1)$ is the probability of type I error one wishes to spend at information time s_1. For a given alpha spending function $\alpha(s)$ and a series of standardized test statistic Z_k, k=1,. . .,K. The corresponding boundaries c_k, k=1,. . .,K are chosen such that under the null hypothesis

$$P(Z_1 < c_1, \ldots, Z_{k-1} < c_{k-1}, Z_k > c_k) = \alpha(\frac{k}{K}) - \alpha(\frac{k-1}{K}).$$

Some commonly used alpha-spending functions are summarized in the Table 1.2.

1.2.4 Sample size re-estimation

In clinical trials, the fact that many parameters, such as assumed treatment effect size, are uncertain will cause the study to be under-powered or over-powered. Assuming a conservative effect size and designing a trial with a larger sample size is one solution. Without a large enough number of sample size, a clinical trial, especially a phase III study design cannot be convincing from a scientific or a financial viewpoint. To ensure a desirable power, sample size re-estimation (SSR) design has been proposed. In SSR design, a sample size based on an guessed effect size is calculated before the study. In an interim analysis, the sample size is re-estimated adaptively based on the accrued data and the target power.

Let us assume a randomized trial with two parallel groups (a test treatment vs. a placebo). Assume that the distribution of the response of the primary endpoint is distributed as a normal distribution. The total sample size required for obtaining a desired power of $1-\beta$ for a two-sided alternative hypothesis can be obtained using the following formula (see, e.g., Chow et al., 2003)

$$N = \frac{4\sigma^2(z_{\alpha/2} + z_\beta)}{\Delta^2}$$

where Δ is the clinically meaningful difference. Usually, σ^2 is unknown and need to be estimated based on previous studies. Let σ^{*2} be the initial guess of the within-group variance for sample size determination before the study. Nevertheless, if the true within-group variance is actually $\sigma^{'2}$, then the sample size to be adjusted to achieve the desired power 1-β at the α level of significance for a two-sided alternative is given by

$$N^{'} = N\frac{\sigma^{'2}}{\sigma^{*2}}$$

Various statistical procedures for sample size re-estimation in group sequential

trial designs are proposed, such as, Cui-Hung-Wang's method (1999), Proschan-Hunsberger's method (1995), and Bauer and Köhne's idea (1994).

In the Cui-Hung-Wang's approach (1999), suppose that it is planned to perform up to K-1 interim analyses and one possible final analysis and that n_k subjects are obtained for each population between the (k-1)th and kth analyses. let N_k be the planned cumulative sample size from stage 1 to stage k, and let $t_k = N_k/N$ be the information fraction or information time at the kth interim analysis. At the end of the Lth interim analysis for specified $L(1 \leq L \leq K-1)$, the adjusted total sample size based on the observed treatment effect Δ_L is

$$M = N(\Delta/\Delta_L)^2. \tag{1.5}$$

Accordingly, the sample size at $(L+j)$th look is

$$M_{L+j} = b(N_{L+j} - N_L) + N_L, \tag{1.6}$$

where $b = (M - N_L)/(N - N_L), j = 1, \ldots, K-L$. They developed a new group sequential test based on the repeated significance test that can be asymptotically expressed as a Brownian motion process. Let $B(t)$ be such a repeated significance test evaluated at the information time $t, 0 \leq t \leq 1$. Let $Z(t) = B(t)/t^{1/2}$. Suppose that the decision to increase the maximum information from one to ω is made at time $t = t_L$ on the basis of the observed value of $Z(t_L)$. Let $c = (\omega - t_L)/(1 - t_L)$. Thus the new test statistic can be constructed as

$$U(t) = Z(t), t \leq t_L, \tag{1.7}$$

and

$$U(t) = Z(t_L)\{\omega(t_L,t)\}^{1/2}+[\{B(c(t-t_L)+t_L)-B(t_L)\}/\{c(t-t_L)\}^{1/2}]\times[1-\omega(t_L,t)]^{1/2}, t_L \leq t \leq 1, \tag{1.8}$$

where $\omega(t_L,t) = t_L/t$. Cui et al. (1999) showed that using U(t) and original boundary from the group sequential trial will not inflate the type I error rate, but gain power substantially.

Cui-Hung-Wang's method has the following advantages. First, the adjustment of sample size is easy. Second, using the same stopping boundaries from the traditional group sequential trial is straightforward. The disadvantages include that (i) this method is somewhat ad hoc, which does not aim a target power, and (ii) Weighting outcomes differently for patients from different stages is difficult to explain clinically.

For a given two-stage design, Proschan and Hunsberger(1995) and Proschan (2005) proposed re-estimating sample size based on the conditional power and offered a new critical value to control the type I error rate.

Chow and Chang (2011) discussed the SSR methods for Bauer-Köhne's (1994) sequential method approach. In the Bauer-Köhne method, let P_1 and P_2 be the p-values for the sub-samples obtained from the first stage and second stage, respectively. Fisher's criterion leads to rejection of H_0 at the end of trial if

$$P_1 P_2 \leq c_\alpha = e^{-\frac{1}{2}\chi^2_{4,1-\alpha}}.$$

Decision rules at the first stage:

$$\begin{cases} P_1 \leq \alpha_1, & \text{Stop trial and reject } H_0, \\ P_1 > \alpha_0, & \text{Stop trial and accept } H_0, \\ \alpha_1 < P_1 \leq \alpha_0, & \text{Continue to the second stage.} \end{cases}$$

For determination of α_1 and α_0, the overall type I error rate is given by

$$\alpha_1 + \int_{\alpha_1}^{\alpha_0} \int_0^{\frac{c_\alpha}{P_1}} dP_2 dP_1 = \alpha_1 + c_\alpha ln \frac{\alpha_0}{\alpha_1} = \alpha.$$

Decision rule at the final stage is given by

$$\begin{cases} P_1 P_2 \leq e^{-\frac{1}{2}\chi^2_{4,1-\alpha}}, & \text{Reject } H_0, \\ Otherwise, & \text{Accept } H_0. \end{cases}$$

Lai (2013) studied the effect of classic Brownian and fractional Brownian motion on the sample size estimation with interim analysis. The fundamental assumptions in the Brownian motion is that the increment of the monitoring statistic would be independent. Nevertheless, this assumption may be violated due to aggregation. The fractional Brownian motion is an extension of the classic Brownian motion, which have a long memory to apply to interim analysis.

1.3 Public health significance

Clinical trials are the gold standard for evaluating new therapies. 'A properly planned and executed clinical trial is the best experimental technique for assessing the effectiveness of an intervention. It also contributes to the identification of possible harms (Friedman et al., 2015).' The clinical trial directly involves human beings and cost a lot. According to the 2015 - 2016 Global Participation in Clinical Trials Report by FDA, 'the country contributing the most clinical trial participants was the United States. Compared to the population of the entire world (7.4 Billion), the US (0.35 Billion) makes up a little more than 4% of the world population.' 'A Phase 2 clinical trial costs from US $7.0 million (cardiovascular) to US $19.6 million(hematology), whereas a Phase 3 clinical

trial costs ranged from US $11.5 million (dermatology) to US $52.9 (pain and anesthesia) on average' (Sertkaya, A. et al., 2016).

But traditional clinical trials may suffer from several flaws, exposing patients to inferior treatments and danger and wasting resources and money. Therefore, there is an urgent need to develop efficient and ethical clinical trial designs and analysis methods. Response-adaptive randomization can achieve different ethical and efficient objectives. Covariate-adaptive randomization is proposed to eliminate selection biases and imbalance of covariates across treatments, leading to better analysis of trial results. Sequential monitoring possesses ethical, administrative and economic advantages. Sample size re-estimation is an useful approach to guarantee the power and success of a trial.

In this dissertation, I study statistical properties of combining sequential monitoring, SSR and adaptive randomization in one clinical trial. The success of the research can lead to a more efficient and ethical trial with effective sample size, saving more patients in the trial and benefiting the general population related to the corresponding treatments.

1.4 Organization of the dissertation

In Chapter 2, I study sequential monitoring of urn models with SSR. In Chapter 3, I study sequential monitoring of CAR with SSR when all the randomization covariates are included in the data analysis. In Chapter 4, I study sequential monitoring of CAR with SSR when a subset of the randomization covariates are included in the data analysis. The conclusions are in Chapter 5, followed by the reference, and the proofs are in the Appendix at the end of the dissertation.

Table 1.1: Unified Formulation for Sequential Design

Single mean	$Z_k = (\overline{X}_k - \mu_0)\sqrt{I_k}$	$I_k = \frac{n_k}{\sigma^2}$
Paired means	$Z_k = \overline{D}_k\sqrt{I_k}$	$I_k = \frac{n_k}{\widetilde{\sigma}^2}$
Two means	$Z_k = (\overline{X}_{Ak} - \overline{X}_{Bk})\sqrt{I_k}$	$I_k = (\frac{\sigma_A^2}{n_{Ak}} + \frac{\sigma_B^2}{n_{Bk}})^{-1}$
One proportion	$Z_k = (\hat{p}_k - p_0)\sqrt{I_k}$	$I_k = \frac{n_k}{\sigma^2}, \sigma^2 = \overline{p}(1-\overline{p})$
Two proportions	$Z_k = (\hat{p}_{Ak} - \hat{p}_{Bk})\sqrt{I_k}$	$I_k = \frac{1}{\sigma^2}(\frac{1}{n_{Ak}} + \frac{1}{n_{Bk}})^{-1}$, $\sigma^2 = \overline{p}(1-\overline{p})$

Table 1.2: $\alpha-$spending functions (Chow and Chang, 2012)

O'Brien-Fleming	$\alpha_1(s) = 2\{1 - \phi(z_{\alpha/2}/\sqrt{s})\}$
Pocock	$\alpha_2(s) = \alpha \log[1 + (e-1)s]$
Lan-DeMets-Kim	$\alpha_3(s) = \alpha s^{\Theta}, \Theta > 0$
Hwang-Shih	$\alpha_4(s) = \alpha[(1 - e^{\zeta s}/(1 - e^{-\zeta})], \zeta \neq 0$

Chapter 2

Sequential monitoring of randomized clinical trials with urn models and sample size re-estimation

Abstract: Clinical trials are usually complex involving multiple competitive objectives such as maximizing the power to detect treatment effects while controlling type I error rate, assigning more patients to better treatment and decreasing the total sample size and cost. Response-adaptive randomization (RAR) procedures have been proposed to achieve these objectives. Sequential monitoring and sample size re-estimation (SSR) are also commonplace in modern clinical trials. In this chapter, I investigate the sequential monitoring of randomized clinical trials with urn models and SSR. To perform sequential monitoring of urn models with SSR, one has to simultaneously address the three sequential procedures (the allocation of patients, the urn compositions and the estimators), and deal with sequential statistics with revised information time due to SSR. Therefore, it is challenging to derive the joint distribution of the sequential statistics, and to control the type I error rate. I overcome these hurdles by employing appropriate framework and SSR

methods, and deriving the asymptotic results for the proposed procedure. Under some regularity conditions, I proved the asymptotic distribution of the proposed sequential statistics follows Brownian motion under null hypothesis. Therefore, traditional critical values for sequential monitoring based on Brownian motion can be used for the proposed procedure to control the type I error rate. I performed simulation studies for three types of urn models, and the results demonstrated that my proposed approaches can control the type I error rate well and also demonstrate the advantages of the proposed methods over traditional designs.

2.1 Introduction

Clinical trials are usually complex involving multiple competitive objectives such as maximizing the power to detect treatment effects , assigning more patients to better treatment and decreasing the total sample size and cost. Practical clinical trials suffer from some inevitable difficulties such as wrong or inaccurate estimate of the required sample size. A variety of adaptive approaches including group sequential monitoring, adaptive randomization, and sample size re-estimation (SSR) have been proposed to solve these problems and achieve ethical and efficient objectives. In this chapter, I study the advantages of the sequential monitoring of clinical trials with randomized urn models and SSR.

It is natural to conduct a sequential analysis in clinical trials where data accumulates sequentially. Jennison and Turnbull (2000) summarized three reasons to perform sequential monitoring in clinical trials. First, it is ethical to monitor progress of the trial to prevent participants from being exposed to unnecessary unsafe, inferior or ineffective treatment regimens. Second, the administrative reason for interim analysis is to ensure that the protocol has been complied. Third, there are obvious economic benefits such as saving cost and time due to possible early stopping.

Sequential monitoring originated from the sequential probability ratio test proposed by Wald (1947). Armitage (1975) introduced sequential monitoring to clinical studies, and his approach was based on a patient-by-patient monitoring. Further, the following three papers are particularly influential and become the foundation of methodological research and basis of practice in clinical trials. Pocock (1977) proposed group sequential monitoring; O'Brien and Fleming (1979) proposed the most popular and commonly used idea of rejection boundaries for sequential monitoring; Lan and DeMets (1983) investigated the alpha spending function that is very flexible and does not require pre-set number of interim analysis and schedule. More details about sequential monitoring can be seen in Jennison and Turnbull (2000), and Whitehead (1997).

Traditional clinical trial designs such as complete randomization and stratified permuted block randomization emphasize equal allocation. For example, Connor et al. (1994) compared the effect of Zidovudine and placebo on reducing maternal-infant HIV transmission with equal allocation. Although the advantages of the new treatment was successfully detected, the randomization was in question. First, we keep assigning patients to the two treatment arms with equal chance even if we have opportunity to detect that the new treatment is probably better during the trial. It is desirable to assign patients to possibly favorable treatment with higher chance, and such strategy potentially increases the enrollment rate. Second, equal allocation has been deemed as the best in terms of power assuming the variances of the two groups are equal, but the assumption may not be true in practice. In order to achieve better ethical and efficient objectives, response-adaptive randomization (RAR) that skews the allocation probability according to the previous treatment assignments and responses has been proposed (Hu and Zhang, 2004). RAR procedure usually consists of three steps: (1) objectives are determined and mathematically formulated ; (2) the optimal allocation proportions which are usually the solutions to the optimization problems formulated in the first step are derived; (3)

appropriate RAR procedures are implemented to target the theoretically derived optimal allocation proportions derived in the second step.

The idea of RAR stemmed from Thompson (1933) and Robbins (1952). Hu and Rosenberger (2003) theoretically proved that RAR can increase the efficiency of clinical trials. Tymofyeyev et al. (2007) established a mathematical framework to derive the optimal allocations. Rosenberger et al. (2001) studied an optimal allocation that minimizes the total number of failures while fixing the power. Ivanova and Rosenberger (2000) showed that an unequal allocation can result in a gain in the power. There are two families of RAR, i.e., doubly adaptive biased coin design (Hu and Zhang, 2004) and urn models. In this chapter, I focused on urn models.

The idea of urn models can be traced back to Pólya's urn model (Eggenberger and Pólya, 1923) and the generalized Friedman's urn model (GFU) by Athreya and Karlin (1968). Zelen (1969) proposed the play-the-winner (PW) rule for clinical trials with binary responses. Wei and Durham (1978) investigated the randomized play-the-winner rule that is the most well-known urn models in clinical trials. Real clinical trials using urn models include Rout et al. (1993), Bartlett et al. (1985) and Tamura et al. (1994). Zhang et al. (2006) proposed a family of sequential estimation-adjusted urn model (SEU) that can target any pre-specified treatment allocation proportion such as Neyman allocation (Neyman, 1934), optimal allocation (Rosenberger et al., 2001) and urn allocation and satisfy various needs. The SEU model contains a variety of urn models such as play-the-winner (PW) rule, randomized play-the-winner (RPW) rule and GFUs as its special cases. I study sequential monitoring of clinical trials with urn models and SSR through SEU model.

Usually, in a clinical trial, the sample size is calculated based on overall assumptions and prior studies with knowledge of similar design conditions. Unfortunately, the prior studies often involve different participating populations, medical practices, etc. As

a result, we may have to modify the sample size to ensure the study power. Wittes and Brittain (1990), Gould (1992), Gould and Shih (1992, 1998), Shih (1992) studied SSR approaches using an internal pilot study; Herson and Wittes (1993) studied SSR approaches for a fixed sample test; Cui et al. (1999) and Denne (1996) studied SSR approaches for group sequential tests.

Despite the numerous advantages of the three adaptive approaches (group sequential monitoring, urn models, and SSR), the research on combining them in one clinical trial is lacking in the literature due to the conceptual and theoretical difficulties. One of the critical statistical problems for all confirmatory clinical trials is the control of the type I error rate. However, sequential monitoring tends to inflate the type I error rate due to multiple hypothesis testing; group sequential monitoring involves correlated sequential statistics at different time points; the treatment assignment probabilities of urn models depend on urn composition, allocation of patients and the sequentially estimated unknown parameters; the responses from urn models depend on all the previous treatment assignments and responses; SSR changes the maximum information and introduces extra dependence between the observed data. To perform sequential monitoring of urn models with SSR, one has to simultaneously address the three sequential procedures (the allocation of patients, the urn compositions and the estimators), and deal with sequential statistics with revised information time due to SSR. Therefore, it is challenging to derive the joint distribution of the sequential statistics, and to control the type I error rate. I overcome these hurdles by employing appropriate framework and SSR methods, and deriving the asymptotic results for the proposed procedure. In my study, I proposed a general framework for sequential monitoring clinical trials using urn models and SSR. I also proposed sequential statistics and proved that its asymptotic distribution is a Brownian motion under null hypothesis. Therefore, traditional critical values for sequential monitoring Brownian motion can be used for the proposed procedure to control the type

I error rate. I performed extensive simulations for three types of urn models, and the results demonstrated that my proposed approaches can control the type I error rate very well.

In Section 2.2, I introduce the notation, framework, the proposed methods, examples under the framework, and theoretical findings. In Section 2.3, I present results from simulations. Conclusions are in Section 2.4.

2.2 Sequential monitoring of SEU model with SSR

2.2.1 Notation and framework

I first offer a general framework for sequential monitoring of SEU model, and incorporate SSR later. Assume the patients sequentially enter the clinical trial comparing two treatments, and the originally planned sample size is n. At the beginning, the urn contains $Y_k(0)$ balls of type $k, k = 1, 2$, and write $\boldsymbol{Y}(0) = (Y_1(0), Y_2(0))$. When the ith patient is ready for randomization, $i = 1, 2, \ldots, n$, a ball, say type k, is randomly drew from the urn, and replaced. Then the ith patient will be allocated to treatment k, and the response $\xi_{i,k}$ will be observed. Additional $d_i(k, g, \xi_{i,k})$ balls of type $g, g = 1, 2$, are added to the urn, where $d_i(k, g, \xi_{i,k})$ is a function of $\xi_{i,k}$. Denote $\boldsymbol{Y}(m) = (Y_1(m), Y_2(m))$ as the urn composition after m patients have been randomly assigned; denote matrix $\boldsymbol{D}_m = (d_m(k, g, \xi_{m,k}), k, g = 1, 2)$ as addition rules; denote $\boldsymbol{X}_m = (X_{m,1}, X_{m,2})$ as the observed result of the mth draw ($X_{m,k} = 1$ if the mth draw is the ball of type k, $k = 1, 2$, $X_{m,k} = 0$ otherwise). Then $\boldsymbol{N}(m) = (N_1(m), N_2(m)) = \sum_{i=1}^{m} \boldsymbol{X}_i$ are the number of patients in the treatments and I have $\boldsymbol{Y}(m) = \boldsymbol{Y}(m-1) + \boldsymbol{X}_m \boldsymbol{D}_m$. Further, I assume that $\boldsymbol{\xi}_m = (\xi_{m,1}, \xi_{m,2})$ are independent and identical distributed with unknown parameter $\boldsymbol{\Theta} = (\theta_1, \theta_2)$. To simplify the notation, I use one-dimensional parameter. It is easy to generalize it to multi-dimensional case. Here, only $\xi_{m,k}$ can be observed if the mth patient

is assigned to treatment k, $k = 1, 2$. Without loss of generality, I assume $\boldsymbol{\Theta} = E[\boldsymbol{\xi}_m]$ since I can transform $\boldsymbol{\xi}_m$ and treat the transformation as responses to make this assumption hold if such transformation exists. Further discussion can be found in Gwise et al. (2008) and Hu and Zhang (2004). Then I can obtain the estimator $\hat{\boldsymbol{\Theta}}(m) = (\hat{\theta}_1(m), \hat{\theta}_2(m))$ after m patients with

$$\hat{\theta}_k(m) = \frac{\sum_{i=1}^{m} X_{i,k}\xi_{i,k} + 1}{N_k(m) + 1}, k = 1, 2,$$

where 1 is added to both the numerator and the denominator to avoid discontinuity and problems caused by the case when no patients are in any certain treatment. Note that both the addition rules $\boldsymbol{D}_m = \boldsymbol{D}(\hat{\boldsymbol{\Theta}}(m-1), \boldsymbol{\xi}_m)$ and the generating matrices $\boldsymbol{H}_m = \boldsymbol{H}(\hat{\boldsymbol{\Theta}}(m-1)) = E[\boldsymbol{D}_m|\mathcal{F}_{m-1}]$ depend on previous responses, where the sigma field $\mathcal{F}_{m-1}$ is generated by $\{\boldsymbol{Y}(0), \boldsymbol{Y}(1), \ldots, \boldsymbol{Y}(m-1), \boldsymbol{X}_1, \ldots, \boldsymbol{X}_{m-1}, \boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{m-1}\}$, which implies that it is a type of RAR design.

Let $\lfloor \cdot \rfloor$ denote the floor function and $t = N/n$ be the *information time* when N is the number of enrolled patients. Accordingly, I have $\boldsymbol{N}(\lfloor nt \rfloor) = (N_1(\lfloor nt \rfloor), N_2(\lfloor nt \rfloor))$, where $N_j(\lfloor nt \rfloor) = \sum_{i=1}^{\lfloor nt \rfloor} X_{i,j}, j = 1, 2$, is the number of patients assigned to treatment j at information time t; $\boldsymbol{Y}(\lfloor nt \rfloor) = (Y_1(\lfloor nt \rfloor), Y_2(\lfloor nt \rfloor))$ is the urn composition at information time t; the estimators are $\hat{\boldsymbol{\Theta}}(\lfloor nt \rfloor) = (\hat{\theta}_1(\lfloor nt \rfloor), \hat{\theta}_2(\lfloor nt \rfloor))$, i.e.,

$$\hat{\theta}_1(\lfloor nt \rfloor) = \frac{\sum_{i=1}^{\lfloor nt \rfloor} X_{i,1}\xi_{i,1} + 1}{N_1(\lfloor nt \rfloor) + 1} \quad \text{and} \quad \hat{\theta}_2(\lfloor nt \rfloor) = \frac{\sum_{i=1}^{\lfloor nt \rfloor} X_{i,2}\xi_{i,2} + 1}{N_2(\lfloor nt \rfloor) + 1} \tag{2.1}$$

In this paper, I perform the following hypothesis testing to compare two treatments in clinical trials:

$$H_0 : h(\theta_1) = h(\theta_2) \text{ versus } H_1 : h(\theta_1) \neq h(\theta_2) \ (or \ h(\theta_1) > h(\theta_2)),$$

where h is a $\Re \to \Re$ function of parameters and assumed to be continuous and twice

differentiable on a small neighborhood of $\theta_i, i = 1, 2$. The following sequential test statistics at time point $t \in (0, 1]$ will be used, i.e.,

$$Z_t\left(\frac{\boldsymbol{N}(\lfloor nt \rfloor)}{\lfloor nt \rfloor}, \hat{\boldsymbol{\Theta}}(\lfloor nt \rfloor)\right) = \frac{h\left(\hat{\theta}_1(\lfloor nt \rfloor)\right) - h\left(\hat{\theta}_2(\lfloor nt \rfloor)\right)}{\sqrt{\hat{Var}\left(h\left(\hat{\theta}_1(\lfloor nt \rfloor)\right)\right) + \hat{Var}\left(h\left(\hat{\theta}_2(\lfloor nt \rfloor)\right)\right)}} \tag{2.2}$$

Assume $\hat{Var}\left(h\left(\hat{\theta}_1\left(\lfloor nt \rfloor\right)\right)\right)$ and $\hat{Var}\left(h\left(\hat{\theta}_2\left(\lfloor nt \rfloor\right)\right)\right)$ are consistent estimators of the variances of $h\left(\hat{\theta}_1\left(\lfloor nt \rfloor\right)\right)$ and $h\left(\hat{\theta}_2\left(\lfloor nt \rfloor\right)\right)$, respectively. We also assume there exist two functions u_1 and u_2 satisfying

$$\lfloor nt \rfloor \hat{Var}\left(h\left(\hat{\theta}_i\left(\lfloor nt \rfloor\right)\right)\right) = u_i\left(\frac{\boldsymbol{N}\left(\lfloor nt \rfloor\right)}{\lfloor nt \rfloor}, \hat{\boldsymbol{\Theta}}\left(\lfloor nt \rfloor\right)\right)(1 + o(1)) \quad \text{a.s. } i = 1, 2.$$

2.2.2 Examples

As a type of RAR design, the SEU model is able to target some pre-specified allocation proportions that are usually derived based on certain optimization criterion. In this chapter, I denote the targeted allocation proportion as $\boldsymbol{v} = (v_1, v_2)$, and details regarding the relationship between $\boldsymbol{v}$ and the generating matrix $\boldsymbol{H}$ can be seen in Zhang et al. (2006). Next, I offer 3 examples to show how to sequentially monitor the SEU model, and the simulations in Section 2.3 are based on the three examples.

Example 1 Assume the responses are binary with success rates p_1 and p_2 for the two treatments under study, and the hypotheses to test are

$$H_0 : p_1 = p_2 \text{ versus } H_1 : p_1 > p_2.$$

The SEU model targeting the following optimal allocation proportion proposed by Rosen-

berger et al. (2001) is used to sequentially assign patients,

$$v_1 = \frac{\sqrt{p_1}}{\sqrt{p_1}+\sqrt{p_2}} \qquad v_2 = \frac{\sqrt{p_2}}{\sqrt{p_1}+\sqrt{p_2}}. \tag{2.3}$$

This optimal allocation is used to minimize the expected total failure number for fixed power. Then the updating rule of balls in the urn and the generating matrice can be derived based on $\boldsymbol{v} = (v_1, v_2)$, where $\boldsymbol{v}$ is the left eigenvector of the limiting generating matrix $\boldsymbol{H}$ with respect to its largest eigenvalue and satisfying $v_1 + v_2 = 1$. For this case, I have

$$\boldsymbol{H} = \begin{pmatrix} \sqrt{p_1} & \sqrt{p_2} \\ \sqrt{p_1} & \sqrt{p_2} \end{pmatrix},$$

and the corresponding addition rule is that $\sqrt{\hat{p}_1(m-1)}$ balls of type 1 and $\sqrt{\hat{p}_2(m-1)}$ balls of type 2 are added to the urn after the mth patient has been randomly assigned.

In this case, $\boldsymbol{\Theta} = (p_1, p_2)$, $h(\theta_j) = \theta_j = p_j, j = 1, 2$, and the sequential statistics $Z_t(\boldsymbol{y}, \boldsymbol{z})$ is a function from $\Re^4$ to $\Re$:

$$Z_t(\boldsymbol{y}, \boldsymbol{z}) = Z_t(y_1, y_2, z_1, z_2) = \frac{z_1 - z_2}{\sqrt{\frac{z_1(1-z_1)}{\lfloor nt \rfloor y_1} + \frac{z_2(1-z_2)}{\lfloor nt \rfloor y_2}}} = \frac{\hat{p}_1(\lfloor nt \rfloor) - \hat{p}_2(\lfloor nt \rfloor)}{\sqrt{\frac{\hat{p}_1(\lfloor nt \rfloor)(1-\hat{p}_1(\lfloor nt \rfloor))}{N_1(\lfloor nt \rfloor)} + \frac{\hat{p}_2(\lfloor nt \rfloor)(1-\hat{p}_2(\lfloor nt \rfloor))}{N_2(\lfloor nt \rfloor)}}},$$

where $\boldsymbol{y} = (N_1(\lfloor nt \rfloor)/\lfloor nt \rfloor, N_2(\lfloor nt \rfloor)/\lfloor nt \rfloor)$ and $\boldsymbol{z} = (\hat{\theta}_1(\lfloor nt \rfloor), \hat{\theta}_2(\lfloor nt \rfloor))$, $h(\hat{\theta}_j(\lfloor nt \rfloor)) = \hat{\theta}_j(\lfloor nt \rfloor) = \hat{p}_j(\lfloor nt \rfloor), j = 1, 2$. I also have

$$\hat{Var}\left(h(\hat{\theta}_j(\lfloor nt \rfloor))\right) = \frac{\hat{p}_j(\lfloor nt \rfloor)(1 - \hat{p}_j(\lfloor nt \rfloor))}{N_j(\lfloor nt \rfloor)}$$

and

$$u_j(\boldsymbol{v}, \boldsymbol{\Theta}) = \frac{p_j(1-p_j)}{v_j}, j = 1, 2.$$

Example 2 (Randomized play-the-winner (RPW) rule) In this example, assume

the responses are binary with success rates p_1 and p_2 for the two treatments under study, and I use SEU model to implement the RPW rule with the targeted urn allocation proportion,

$$v_1 = \frac{q_2}{q_1 + q_2} \qquad v_2 = \frac{q_1}{q_1 + q_2}, \tag{2.4}$$

where $q_j = 1 - p_j, j = 1, 2$. The hypotheses to test are

$$H_0 : p_1 = p_2 \text{ versus } H_1 : p_1 > p_2.$$

The addition rule is that one ball of the same type is added to the urn if the response is success and one ball of the opposite type is added to the urn if the response is failure. So I have

$$\boldsymbol{D}_n = \begin{pmatrix} \xi_{n,1} & 1 - \xi_{n,1} \\ 1 - \xi_{n,2} & \xi_{n,2} \end{pmatrix},$$

and

$$\boldsymbol{H} = \begin{pmatrix} p_1 & q_1 \\ q_2 & p_2 \end{pmatrix}.$$

The sequential statistics are the same as in Example 1.

Example 3. Assume the responses of the two treatments follow normal responses $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$, respectively. The hypothesis are

$$H_0 : \mu_1 = \mu_2 \text{ versus } H_1 : \mu_1 > \mu_2.$$

The SEU model targeting the following Neyman allocation (Neyman, 1934) is used to sequentially assign patients:

$$v_1 = \frac{\sigma_1}{\sigma_1 + \sigma_2} \qquad v_2 = \frac{\sigma_2}{\sigma_1 + \sigma_2}, \tag{2.5}$$

Neyman allocation is used to maximize the power. Based on Neyman allocation, I can derive the generating matrix as follows,

$$\boldsymbol{H} = \begin{pmatrix} \frac{\sigma_1}{\sigma_1+\sigma_2} & \frac{\sigma_2}{\sigma_1+\sigma_2} \\ \frac{\sigma_1}{\sigma_1+\sigma_2} & \frac{\sigma_2}{\sigma_1+\sigma_2} \end{pmatrix},$$

and the addition rule is that $\frac{\hat{\sigma}_1(m-1)}{\hat{\sigma}_1(m-1)+\hat{\sigma}_2(m-1)}$ balls of type 1 and $\frac{\hat{\sigma}_2(m-1)}{\hat{\sigma}_1(m-1)+\hat{\sigma}_2(m-1)}$ balls of type 2 are added to the urn after the mth patient has been randomly assigned.

The test statistics at time t is then

$$Z_t = \frac{\hat{\mu}_1(\lfloor nt \rfloor) - \hat{\mu}_2(\lfloor nt \rfloor)}{\sqrt{\frac{\hat{\sigma}_1(\lfloor nt \rfloor)^2}{N_1(\lfloor nt \rfloor)} + \frac{\hat{\sigma}_2(\lfloor nt \rfloor)^2}{N_2(\lfloor nt \rfloor)}}}.$$

2.2.3 Incorporation of sample size re-estimation

Next, I implement sample size re-estimation in the above procedure of sequential monitoring of SEU models. In this chapter, I assume non-decrease of sample size as recommended by (FDA, 2010). Suppose I have K interim analyses at information time points $t_1, \ldots, t_L, \ldots, t_K$, and I implement SSR at the end of the Lth interim analysis ($L < K$) based on the observed data using the method in Cui et al. (1999). Define the treatment effect (Δ) as

$$\Delta = \frac{\mu_1 - \mu_2}{\sqrt{\frac{\sigma_1^2}{v_1} + \frac{\sigma_2^2}{v_2}}}$$

for normal distribution and

$$\Delta = \frac{p_1 - p_2}{\sqrt{\frac{p_1 q_1}{v_1} + \frac{p_2 q_2}{v_2}}}$$

for binary responses. Because $B_t - \sqrt{n}\Delta t$ is asymptotically standard Brownian Motion according to Zhu and Hu (2012), the conditional power, CP_L, given the observed treatment effect (Δ) and the test statistics (Z_t) at time $t = t_L$ with sample size N can be

calculated in this way

$$
\begin{aligned}
CP_L &= P(Z_1 > C | Z_t, \Delta) \\
&= P\left(Z_1 - \sqrt{n}\Delta > C - \sqrt{n}\Delta | Z_t, \Delta\right) \\
&= P\left(Z_1 - \sqrt{n}\Delta - (\sqrt{t}Z_t - \sqrt{n}\Delta t) > C - \sqrt{n}\Delta - (\sqrt{t}Z_t - \sqrt{n}\Delta t) | Z_t, \Delta\right) \\
&= P\left(\frac{Z_1 - \sqrt{n}\Delta - (\sqrt{t}Z_t - \sqrt{n}\Delta t)}{\sqrt{1-t}} > \frac{C - \sqrt{n}\Delta - (\sqrt{t}Z_t - \sqrt{n}\Delta t)}{\sqrt{1-t}} | Z_t, \Delta\right) \\
&= 1 - \Phi\left(\frac{C - \sqrt{t}Z_t - \sqrt{n}\Delta(1-t)}{\sqrt{1-t}}\right)
\end{aligned}
$$

where $\Phi(\cdot)$ is the CDF of the standard normal distribution, and C is the final critical value at the end of the trial. Specifically, I re-estimate the sample size as follows:

(1) Estimate the treatment effect (Δ) and calculate the test statistics (Z_t) at time $t = t_L$ based on observed sample size N.

(2) If the conditional power, CP_L, calculated by plugging in the estimated treatment effect and observed test statistics from step (1) for originally planned sample size n is not less than the desirable level cp_1, then no SSR will be implemented. Otherwise, if the CP_L is more than 0.01, search n^* that satisfies $CP_L = cp_1$.

(3) Then I 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 I can use the following new sequential statistics to perform sequential monitoring

$$
U_t = \begin{cases} Z_t, & \text{if } t \leq t_L; \\ [w(t_L,t)]^{1/2} \times Z_{t_L} + [1 - w(t_L,t)]^{1/2} \times & \\ \quad \{[B(b(t-t_L)+t_L) - B(t_L)]/[b(t-t_L)]^{1/2}\}, & \text{if } t > t_L, \end{cases} \tag{2.6}
$$

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

2.2.4 Asymptotic results

We need the following assumptions for responses $\boldsymbol{\xi}_n$, addition rules $\boldsymbol{D}_n = \boldsymbol{D}(\hat{\boldsymbol{\Theta}}(n-1), \boldsymbol{\xi}_n)$ and the function $\boldsymbol{H}(\boldsymbol{x})$.

(A1) There exists a constant $\gamma > 0$ such that $\boldsymbol{H}\mathbf{1}' = \gamma\mathbf{1}'$ and $\mathbf{1} = (1, \ldots, 1)$. In addition, $\boldsymbol{H}$ has the following Jordan decomposition:

$$\boldsymbol{T}^{-1}\boldsymbol{H}\boldsymbol{T} = \gamma \text{diag}\,[1, \boldsymbol{J}_2, \ldots, \boldsymbol{J}_s]$$

where $\boldsymbol{J}s$ is a $\nu_t \times \nu_t$ matrix, given by

$$\boldsymbol{J}_t = \begin{pmatrix} \lambda_t & 1 & 0 & \ldots & 0 \\ 0 & \lambda_t & 1 & \ldots & 0 \\ 0 & 0 & \lambda_t & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & \lambda_t \end{pmatrix},$$

and $\boldsymbol{T}$ and $\boldsymbol{J}_t$ are functions of $\boldsymbol{\Theta}$.

(A2) $E\|\boldsymbol{\xi}_1\|^r < \infty$ for some $r > 2$.

(A3) The addition rules $\boldsymbol{D}_n \geq 0$ are bounded.

(A4) $\boldsymbol{H}(x)$ is twice differentiable.

Theorem 2.1. *Let $B_t^U = \sqrt{t}U_t$. If Assumptions (A1)-(A4) are satisfied, then under H_0, B_t^U converges to a standard Brownian motion in distribution. The sequential statistics $\{(U_{t_1}, \ldots, U_{t_K}), 0 \leq t_1 \leq t_2 \leq \ldots \leq t_K \leq 1\}$ follows the asymptotic canonical joint distribution defined in Jennison and Turnbull (2000): under H_0,*

(i) $\{U_{t_1}, \ldots, U_{t_K}\}$ is multivariate normal;

(ii) $EU_{t_i} = 0$;

and (iii) $Cov(U_{t_i}, U_{t_j}) = \sqrt{\lfloor nt_i \rfloor / \lfloor nt_j \rfloor},\ 0 \le t_i \le t_j \le 1.$

The proof is given in the Appendix.

This theorem reveals the most fundamental properties for the proposed method, i.e., the asymptotic joint distribution of the sequential statistics. Therefore, a variety of future research and methods can be performed based on this result, among which the control of the type I error rate is the focus of this chapter. Since the asymptotic joint distribution of the sequential statistics is the asymptotic canonical joint distribution defined in Jennison and Turnbull (2000), all the methods based on this distribution in that book and in other papers such as Pocock's test, O'Brien and Fleming's test, the tests of Wang and Tsiatis (1987), the tests of Haybittle (1971) and Peto et al. (1976), the equivalence test, spending functions, stochastic curtailment, and repeated confidence intervals can be used to control the type I error rate for this procedure and to provide important information for DSMB to make decision about whether to continue the trial. In this chapter, I use the alpha spending function mimicking the O'Brien Fleming boundaries as follows,

$$\alpha_{OBF}(t) = 2\left(1 - \Phi\left(z_{\alpha/2}/\sqrt{t}\right)\right).$$

If I perform the sequential monitoring at information time $t_1 = 0.2, t_2 = 0.5$, and $t_3 = 1$, the corresponding boundaries are $C_1 = 4.877$, $C_2 = 2.963$, $C_3 = 1.969$ (Proschan et al., 2006).

2.3 Numerical and simulation studies

In this section, I study the finite-sample properties of my proposed methods using the three SEU models in Example 1-3, and compare the SEU models with complete randomization. Assume that the originally planned sample size is $n = 500$ with three

interim looks at information time $t_1 = 0.2$ $(n_1 = 100), t_2 = 0.5$ $(n_2 = 250)$, and $t_3 = 1$ $(n = 500)$. The corresponding O'Brien-Fleming-like spending function boundaries are $C_1 = 4.877$, $C_2 = 2.963$, $C_3 = 1.969$. I 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. In this case, $w(t_2, t_3) = 0.5$ and $b_{\max} = 2$. The datasets are generated based on different parameter combinations shown in the tables. All the results are based on $10,000$ replications.

Table 2.1 summarizes the results of the SEU model in Example 1. The initial urn composition is $\boldsymbol{Y}(0) = (5, 5)$, and the randomization procedure will follow the rule of urn models explained before. If I decide to continue the trial after the second interim look as described in Section 2.2.3, I calculate the conditional power based on the observed data. If the conditional power is less than 0.9, I increase the sample size to make the conditional power to be 0.9. I report the type I error rate (α) (the proportion of the number of rejections of H_0 out of 10,000 replications, and the intended value is 0.025) and the average and standard deviation of the following values out of $10,000$ replications: actual allocation proportion in treatment 1 ($\hat{\rho}_1 = N_1/(N_1 + N_2)$), urn compositions represented by the proportion of balls of type 1 ($\hat{Urn}_1 = Y_1/(Y_1 + Y_2)$), total sample size (SS), total failure number (failure) and failure rate (failrate) considering the total sample sizes are different. I found that my proposed method can control the type I error rate very well. From the results that both $\hat{\rho}_1$ and $\hat{Urn}_1$ are close to 0.5, I can see that my method converges very well. My method does not increase the total sample size.

Table 2.2 reports the empirical power and the average and standard deviation of the following values out of $10,000$ replications: actual allocation proportion in treatment 1 ($\hat{\rho}_1$), urn compositions represented by the proportion of balls of type 1 ($\hat{Urn}_1$), total sample size (SS), total failure number (failure) and failure rate (failrate). I find that my proposed method can assign more patients to the better treatment and lead to fewer

failures while controlling the power at the same level as complete randomization (CR), which is consistent with the objective of the optimal allocations (2.3).

In Table 2.3, I study the SEU model in example 2. Other settings are the same as in Table 1. I obtain similar conclusions as in Table 2.1. In Tables 2.1 and 2.3, since there is no treatment effects, the targeted allocation proportions for both SEU models are equal allocation. Therefore, the SEU models perform equivalently to complete randomization in terms of allocation proportion and number of failures. In Tables 2.1 and 2.3, I focus on the results of type I error rate. In Table 2.4, I study the performance of RPW rule under H_1. I find that my proposed method can assign more patients to the better treatment and lead to fewer failures, which is consistent with the objective of the urn allocations (2.4).

In Table 2.5, I study the performance of the SEU model in Example 3 under H_0. In order to get the initial estimate of unknown parameter to update the urn, I randomly assign 20 patients to the two treatments equally. I found that my proposed method can control the type I error rate very well. From the results of $\hat{\rho}_1$ and $\hat{Urn}_1$, I can see that my method converges very well. In Table 2.6, I study the performance of the SEU model in Example 3 under H_1. I can see that the SEU model targeting the Neyman allocation can increase the power.

2.4 Conclusion

RAR designs have been well-accepted to better achieve various ethical and efficient objectives. In order to promote its application in real clinical trials, it is necessary to study statistical properties of combining RAR and the commonly used procedures in clinical trials, such as sequential monitoring and sample size re-estimation. This chapter addressed this problem using urn models. I established asymptotic results of the proposed method and performed comprehensive simulations to demonstrate that I can control the

type I error rate with advantages of assigning more patients to the better treatments, increasing the power and stopping the trial earlier if necessary.

In this chapter, I used alpha spending function to control the type I error rate. Other methods such as the optimal spending functions in Anderson (2007) and the beta spending functions in DeMets (2006) can be investigated. I assumed that the responses are immediately available, which is not always true in real clinical trials. However, there is no difficulty in incorporating delayed responses into the RAR procedure (Hu and Rosenberger, 2006). We can always update the parameter estimators with collected data. It is worth noting that Bai et al. (2008) and Hu and Zhang (2004) showed that the asymptotic results for GFU were not be affected if the response time is reasonably large compared to the entry time intervals. Hu et al. (2008) studied the effect of delayed responses on DBCD. I leave these for future research.

Table 2.1: Performance of different designs under H_0 for the scenario of Example 1

(p_1, p_2)	Design	α	$\hat{\rho}_1$	$\hat{Urn}_1$	SS	failure	failrate
$(0.2, 0.2)$	SEU1	0.025	0.500(0.038)	0.501(0.026)	594(121)	475(98)	0.800(0.017)
$(0.2, 0.2)$	CR	0.026	0.500(0.021)	NA	594(122)	475(98)	0.800(0.017)
$(0.3, 0.3)$	SEU1	0.028	0.500(0.034)	0.501(0.021)	593(122)	416(86)	0.700(0.019)
$(0.3, 0.3)$	CR	0.026	0.500(0.021)	NA	595(122)	416(86)	0.700(0.019)
$(0.5, 0.5)$	SEU1	0.026	0.501(0.028)	0.501(0.014)	595(122)	298(62)	0.500(0.021)
$(0.5, 0.5)$	CR	0.028	0.500(0.021)	NA	597(123)	298(63)	0.500(0.021)
$(0.7, 0.7)$	SEU1	0.027	0.500(0.025)	0.501(0.009)	594(123)	178(38)	0.300(0.019)
$(0.7, 0.7)$	CR	0.030	0.500(0.021)	NA	593(122)	178(38)	0.300(0.019)
$(0.8, 0.8)$	SEU1	0.026	0.500(0.023)	0.500(0.007)	596(122)	119(26)	0.200(0.017)
$(0.8, 0.8)$	CR	0.027	0.500(0.021)	NA	592(122)	118(26)	0.200(0.017)

Note: p_1, p_2 are the success rates of the binary response for the two treatments; α is the proportion of the number of rejections of H_0 out of 10,000 replications and the intended value is 0.025; actual allocation proportion in treatment 1 ($\hat{\rho}_1$), urn compositions represented by the proportion of balls of type 1 ($\hat{Urn}_1$), total sample size (SS), total failure number (failure) and failure rate (failrate) are the average and standard deviation of the values out of $10,000$ replications.

Table 2.2: Performance of different designs under H_1 for the scenario of Example 1

(p_1, p_2)	Design	Power	$\hat{\rho}_1$	$\hat{Urn}_1$	SS	failure	failrate
$(0.2, 0.325)$	SEU1	0.953	0.453(0.037)	0.443(0.025)	526(187)	385(138)	0.731(0.021)
$(0.2, 0.325)$	CR	0.953	0.500(0.024)	NA	529(188)	390(139)	0.737(0.021)
$(0.3, 0.425)$	SEU1	0.917	0.464(0.033)	0.458(0.020)	555(184)	352(118)	0.633(0.022)
$(0.3, 0.425)$	CR	0.921	0.500(0.023)	NA	557(184)	355(118)	0.637(0.022)
$(0.75, 0.875)$	SEU1	0.982	0.483(0.027)	0.481(0.008)	476(189)	89(36)	0.186(0.019)
$(0.75, 0.875)$	CR	0.982	0.501(0.025)	NA	478(191)	90(36)	0.188(0.020)

Table 2.3: Performance of different designs under H_0 for the scenario of Example 2

(p_1, p_2)	Design	α	$\hat{\rho}_1$	$\hat{Urn}_1$	SS	failure	failrate
$(0.1, 0.1)$	SEU2	0.026	0.500(0.015)	0.501(0.013)	595(122)	536(110)	0.900(0.013)
$(0.1, 0.1)$	CR	0.024	0.500(0.021)	NA	596(122)	536(110)	0.900(0.013)
$(0.3, 0.3)$	SEU2	0.026	0.501(0.023)	0.501(0.016)	594(122)	416(86)	0.700(0.019)
$(0.3, 0.3)$	CR	0.026	0.500(0.021)	NA	595(122)	416(86)	0.700(0.019)
$(0.6, 0.6)$	SEU2	0.025	0.502(0.041)	0.502(0.026)	595(122)	238(50)	0.400(0.021)
$(0.6, 0.6)$	CR	0.032	0.500(0.021)	NA	595(123)	238(50)	0.400(0.020)
$(0.7, 0.7)$	SEU2	0.025	0.501(0.053)	0.502(0.034)	594(121)	178(38)	0.300(0.019)
$(0.7, 0.7)$	CR	0.030	0.500(0.021)	NA	593(122)	178(38)	0.300(0.019)
$(0.8, 0.8)$	SEU2	0.022	0.499(0.072)	0.501(0.051)	595(122)	119(26)	0.200(0.017)
$(0.8, 0.8)$	CR	0.027	0.500(0.021)	NA	592(122)	118(26)	0.200(0.017)

Table 2.4: Performance of different designs under H_1 for the scenario of Example 2

(p_1, p_2)	Design	Power	$\hat{\rho}_1$	$\hat{Urn}_1$	SS	failure	failrate
$(0.2, 0.325)$	SEU2	0.951	0.458(0.023)	0.453(0.018)	530(186)	388(138)	0.732(0.021)
$(0.2, 0.325)$	CR	0.953	0.500(0.024)	NA	529(188)	390(139)	0.737(0.021)
$(0.3, 0.425)$	SEU2	0.913	0.452(0.028)	0.446(0.020)	559(182)	353(116)	0.631(0.023)
$(0.3, 0.425)$	CR	0.921	0.500(0.023)	NA	557(184)	355(118)	0.637(0.022)
$(0.7, 0.825)$	SEU2	0.963	0.404(0.064)	0.385(0.044)	533(181)	120(43)	0.226(0.020)
$(0.7, 0.825)$	CR	0.964	0.500(0.024)	NA	513(190)	122(46)	0.238(0.020)

Table 2.5: Performance of different designs under H_0 for the scenario of Example 3

$(\mu_1, \mu_2, \sigma_1, \sigma_2)$	Design	α	$\hat{\rho}_1$	$\hat{Urn}_1$	SS
$(0, 0, 1, 1)$	SEU3	0.024	0.500(0.034)	0.500(0.020)	595(122)
$(0, 0, 1, 1)$	CR	0.027	0.500(0.021)	NA	574(122)
$(1, 1, 1, 1)$	SEU3	0.025	0.500(0.033)	0.500(0.020)	596(123)
$(1, 1, 1, 1)$	CR	0.025	0.500(0.021)	NA	574(122)
$(0, 0, 0.5, 0.5)$	SEU3	0.025	0.500(0.033)	0.500(0.020)	592(122)
$(0, 0, 0.5, 0.5)$	CR	0.024	0.500(0.021)	NA	572(122)
$(5, 5, 2, 2)$	SEU3	0.026	0.500(0.034)	0.500(0.020)	593(122)
$(5, 5, 2, 2)$	CR	0.027	0.500(0.021)	NA	573(122)

Table 2.6: Performance of different designs under H_1 for the scenario of Example 3

$(\mu_1, \mu_2, \sigma_1, \sigma_2)$	Design	Power	$\hat{\rho}_1$	$\hat{Urn}_1$	SS
$(1, 1.4, 1, 2)$	SEU3	0.923	0.353(0.034)	0.336(0.020)	552(184)
$(1, 1.4, 1, 2)$	CR	0.889	0.500(0.023)	NA	554(176)
$(1, 1.4, 1, 1.5)$	SEU3	0.981	0.414(0.036)	0.403(0.022)	483(190)
$(1, 1.4, 1, 1.5)$	CR	0.974	0.500(0.025)	NA	483(188)
$(1, 1.3, 1, 2)$	SEU3	0.735	0.352(0.032)	0.336(0.019)	617(159)
$(1, 1.3, 1, 2)$	CR	0.663	0.500(0.022)	NA	603(153)
$(1, 1.3, 1, 1.5)$	SEU3	0.867	0.412(0.034)	0.402(0.021)	576(177)
$(1, 1.3, 1, 1.5)$	CR	0.839	0.500(0.023)	NA	570(169)
$(1, 1.5, 1, 2)$	SEU3	0.987	0.357(0.036)	0.337(0.022)	468(192)
$(1, 1.5, 1, 2)$	CR	0.974	0.499(0.025)	NA	481(188)

Chapter 3

Sequential monitoring of randomized clinical trials with CAR and SSR-All randomization covariates are included in the data analysis

Abstract: Clinical trials are usually complex and involve multiple covariates of interest. Therefore, incorporating covariates into randomization design is of special importance. In particular, it is well accepted that the balance of treatment allocation among subgroups defined by covariates is critical in evaluating treatment effects without bias. Covariate-adaptive randomization (CAR) procedures have been proposed to achieve this aim. Sequential monitoring and sample size re-estimation are also commonly used in managing clinical trials. In this chapter, I conduct theoretical and simulation study on the sequential monitoring of CAR with sample size re-estimation (SSR). It is worth noting that all the three procedures cause complex interdependence among responses, treatment assignments, covariates, and sequential statistics. I overcame these difficul-

ties, and derived the asymptotic distribution of the proposed sequential statistics and evaluated the type I error rate via simulations.

3.1 Introduction

It is well accepted that the balance of treatment allocation among subgroups defined by covariates is critical to properly assess the treatment effects in clinical trials. Covariate-adaptive randomization (CAR) procedures sequentially assign the patients based on previous assignments and covariates, and the current covariate profile in order to achieve this aim and increases the credibility of a trial (Rosenberger and Lachin, 2015). Stratified permuted block (SPB) randomization is the most efficient way when there are a small number of covariates and small numbers of levels within each covariate (Zelen, 1974). SPB employs permuted block randomization separately within each stratum formed by crossing of covariates levels. However, when there are a larger number of covariates or many levels within certain covariates, the number of patients belonging to each stratum is typically very small, and SPB will work more like complete randomization. As a result, minimization (Taves, 1974) has been proposed to achieve allocation balance on covariate margins, instead of within strata. Pocock and Simon's design has been described in Chapter 1 of the dissertation. Other research on CAR is in Nordle and Brantmark (1977), Wei (1978), Signorini et al. (1993), Heritier et al. (2005), and Hu and Hu (2012). CAR has been widely acknowledged to be able to achieve the balance of covariates across treatments (Rosenberger and Lachin, 2015). However, it raised concerns about its impact on statistical inference due to the complicated dependence among covariates, treatment assignments and responses and the discreteness of the allocation function.

The history of general sequential monitoring and sample size re-estimation has been offered in Chapter 2. For this current chapter, it is worth noting that Jennison

and Turnbull (1997) discussed group sequential analysis methods incorporating covariate information through linear models, general parametric regression models and survival models. However, they did not take into account the problems caused by CAR and the scenario where not all the randomization covariates were included in the data analysis.

In this chapter, I formulated a general framework for sequential monitoring clinical trials using CAR design, linear regression models with all the randomization covariates for analysis and SSR procedure. In the next chapter, I study the sequential monitoring of clinical trials with the CAR design, linear regression models with a subset of the randomization covariates for analysis and SSR procedure. I defined sequential statistics and derived its asymptotic distribution to be a Brownian motion under null hypothesis. Therefore, classic Brownian motion critical values for sequential monitoring can be used for the proposed procedure to control the type I error rate. I performed extensive simulations and the results demonstrated that my proposed approaches can control the type I error rate well.

In Section 2.2, I introduce the notation, framework, my proposed methods, and theoretical findings. In Section 2.3, I offer results from simulation results. Conclusions are in Section 2.4.

3.2 Sequential monitoring of CAR with SSR when all the randomization covariates are in the data analysis

3.2.1 Framework

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 $(X_1, \ldots, X_p)$ are used to implement CAR and 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 their expectations are all 0 without loss of generality, i.e., $E(X_{ik}) = 0, i = 1, \ldots, n, k = 1, \ldots, p$. In addition, the errors are assumed to be independent. Assume that the ith subject's response Y_i follows the linear model:

$$Y_i = \mu_1 T_i + \mu_2(1 - T_i) + X_{i1}\beta_1 + \ldots + X_{ip}\beta_p + \epsilon_i, \tag{3.1}$$

where $\boldsymbol{\mu} = (\mu_1, \mu_2)^T$ is the treatment effect vector for treatments 1 and 2 respectively, $(\beta_1, \ldots, \beta_p)$ are unknown parameters for covariate effects, and the ϵ_i are independent errors with mean 0 and variance σ^2. Here, we do not have to assume the errors follow normal distribution. We write $\boldsymbol{\eta} = (\mu_1, \mu_2, \beta_1, \ldots, \beta_p)^T$, $\boldsymbol{T}(n) = (T_1, \ldots, T_n)^T$, $\boldsymbol{Y}(n) = (Y_1, \ldots, Y_n)^T$, $\boldsymbol{\epsilon}(n) = (\epsilon_1, \ldots, \epsilon_n)^T$ and

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

So we have

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\eta} + \epsilon.$$

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

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

where $C = \{l : \text{index of continuous covariates among } X_l, l = 1, \ldots, p\}$ and $d_j(\cdot)$ is the discrete function.

In this chapter, I perform the following hypothesis testing to compare two treatments in clinical trials:

$$H_0 : \mu_1 = \mu_2 \text{ versus } \mu_1 \neq \mu_2. \tag{3.2}$$

Let $\lfloor \cdot \rfloor$ denote the floor function and $t = N/n$ be the *information time* when N is the number of enrolled patients. A widely used test statistic including all the randomization covariates in the data analysis to test the hypothesis (3.2) at time point $t \in (0, 1]$ is

$$Z_t = \frac{L\hat{\boldsymbol{\eta}}(t)}{\sqrt{\hat{\sigma}(t)^2 L(\boldsymbol{X}(\lfloor nt \rfloor)^T \boldsymbol{X}(\lfloor nt \rfloor))^{-1} L^T}}, \tag{3.3}$$

where $L = (1, -1, 0, \ldots, 0)$, $\hat{\boldsymbol{\eta}}(t) = (\boldsymbol{X}(\lfloor nt \rfloor)^T \boldsymbol{X}(\lfloor nt \rfloor))^{-1} \boldsymbol{X}(\lfloor nt \rfloor)^T \boldsymbol{Y}(\lfloor nt \rfloor)$, $\hat{\sigma}(t)^2 = [\boldsymbol{Y}(\lfloor nt \rfloor) - \boldsymbol{X}(\lfloor nt \rfloor)\hat{\boldsymbol{\eta}}(t)]^T [\boldsymbol{Y}(\lfloor nt \rfloor) - \boldsymbol{X}(\lfloor nt \rfloor)\hat{\boldsymbol{\eta}}(t)]/(\lfloor nt \rfloor - p - 2)$. These sequential statistics (3.3) are the commonly used ones including t-test statistic as a special case when no covariates are included in the model.

3.2.2 Incorporation of sample size re-estimation

We implement SSR in the same way as Section 2.2.3. Note that in this chapter, I are discussing two-sided hypothesis testing while one-sided hypothesis testing was studied in Chapter 2. The conditional power for the two-sided hypothesis testing can be obtained

as follows. According to Zhu and Hu (2018), $B_t - \sqrt{n}\Delta t$ is asymptotically standard Brownian motion, where

$$\Delta = \frac{\mu_1 - \mu_2}{\sqrt{\frac{\sigma_1^2}{v_1} + \frac{\sigma_2^2}{v_2}}}$$

for normal distribution and

$$\Delta = \frac{p_1 - p_2}{\sqrt{\frac{p_1 q_1}{v_1} + \frac{p_2 q_2}{v_2}}}$$

for binary responses. Therefore, I have

$$\begin{aligned}
&CP_L = P(|Z_1| > C | Z_t, \Delta) \\
&= P(Z_1 > C \text{ or } Z_1 < -C | Z_t, \Delta) \\
&= P(Z_1 - \sqrt{n}\Delta > C - \sqrt{n}\Delta \text{ or } Z_1 - \sqrt{n}\Delta < -C - \sqrt{n}\Delta | Z_t, \Delta) \\
&= P(Z_1 - \sqrt{n}\Delta - (\sqrt{t}Z_t - \sqrt{n}\Delta t) > C - \sqrt{n}\Delta - (\sqrt{t}Z_t - \sqrt{n}\Delta t) | Z_t, \Delta) \\
&+ P(Z_1 - \sqrt{n}\Delta - (\sqrt{t}Z_t - \sqrt{n}\Delta t) < -C - \sqrt{n}\Delta - (\sqrt{t}Z_t - \sqrt{n}\Delta t) | Z_t, \Delta) \\
&= P\left(\frac{Z_1 - \sqrt{n}\Delta - (\sqrt{t}Z_t - \sqrt{n}\Delta t)}{\sqrt{1-t}} > \frac{C - \sqrt{n}\Delta - (\sqrt{t}Z_t - \sqrt{n}\Delta t)}{\sqrt{1-t}} | Z_t, \Delta\right) \\
&+ P\left(\frac{Z_1 - \sqrt{n}\Delta - (\sqrt{t}Z_t - \sqrt{n}\Delta t)}{\sqrt{1-t}} < \frac{-C - \sqrt{n}\Delta - (\sqrt{t}Z_t - \sqrt{n}\Delta t)}{\sqrt{1-t}} | Z_t, \Delta\right) \\
&= 1 - \Phi\left(\frac{C - \sqrt{t}Z_t - \sqrt{n}\Delta(1-t)}{\sqrt{1-t}}\right) + \Phi\left(\frac{-C - \sqrt{t}Z_t - \sqrt{n}\Delta(1-t)}{\sqrt{1-t}}\right)
\end{aligned}$$

3.2.3 Asymptotic results

We need the following notations to formulate the main theorem in this chapter. Suppose $\tilde{X}_k$ has s_k levels, and let $\boldsymbol{W}_i = (x_{i1}^{c_1}, \ldots, x_{ip}^{c_p})$ represents the ith subject's covariate profile if $\tilde{X}_{ik}$ is at level $x_{ik}^{c_k}, k = 1, ..., p$. 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 with respect to the level $x_k^{c_k}$ of covariate $\tilde{X}_k$; let $DIF_n(c_1, \ldots, c_p)$ be the difference in patient numbers in the stratum containing the subjects with covariates $(x_1^{c_1}, \ldots, x_p^{c_p})$.

Theorem 3.1. *Let* $B_t^U = \sqrt{t}U_t$. *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^U *is asymptotically a standard Brownian motion in distribution. The sequential statistics* $\{(U_{t_1}, \ldots, U_{t_K}), 0 \leq t_1 \leq t_2 \leq ... \leq t_K \leq 1\}$ *has the asymptotic canonical joint distribution defined in Jennison and Turnbull (2000), i.e., under* H_0,

(i) $\{U_{t_1}, ..., U_{t_K}\}$ *follows multivariate normal distribution;*

(ii) $EU_{t_i} = 0$;

(iii) $Cov(U_{t_i}, U_{t_j}) = \sqrt{[nt_i]/[nt_j]},\ 0 \leq t_i \leq t_j \leq 1.$

The proof is given in the Appendix.

This theorem reveals the most fundamental properties for the proposed method, i.e., the asymptotic joint distribution of the sequential statistics. Therefore, a variety of future research and methods as introduced in Chapter 2 can be performed based on this result, among which the control of the type I error rate is the focus of this chapter. I also note that the conditions, $DIF_n = O_p(1)$ and $DIF_n^X(k; c_k) = O_p(1), k = 1, \ldots, p$, hold for a variety of CAR procedures including as stratified permuted block randomization and Pocock ans Simons's design.

3.3 Numerical and simulation studies

In this section, I study the finite-sample properties of the proposed procedure. For all the tables, suppose originally planned 500 patients sequentially enter a clinical trial, and the responses follow

$$Y_i = \mu_1 T_i + \mu_2(1 - T_i) + X_{i1}\beta_1 + X_{i2}\beta_2 + \epsilon_i, i = 1, \ldots, 500, \tag{3.4}$$

where $(\mu_1, \mu_2, \beta_1, \beta_2)$ are unknown parameters, and ϵ_i are independent errors from the

normal distribution $N(0,1)$. In different tables, I compare the stratified permuted block randomization (SPB), Pocock and Simon's procedure (PS) and complete randomization (CR). The CAR designs will be applied with respect to both X_1 and X_2, and different distributions of these two covariates will be considered. In this chapter, the sequential data analysis are all based on the model (3.4). Equivalently, it can be written as

$$Y_i = \beta_0 + \beta_T T_i + X_{i1}\beta_1 + X_{i2}\beta_2 + \epsilon_i, i = 1, \ldots, 500, \tag{3.5}$$

that is, all the randomization covariates are used in the data analysis. I 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. In this case, $w(t_2, t_3) = 0.5$ and $b_{\max} = 2$. All the results are based on 10,000 replications.

In Table 3.1, I report results for SPB and complete randomization when both X_1 and X_2 are binary covariates with success rates of p_1 and p_2, respectively. I offer results for type I error rate (α) (the proportion of the number of rejections of H_0 out of 10,000 replications, and the intended value is 0.05), average and standard deviation of the following values out of 10,000 replications: estimates of β_1, β_2 and β_T. I can see that my method can control the type I error rate very well, and estimate the parameters very accurately.

In Table 3.2, I report results for SPB and complete randomization when both X_1 and X_2 follow standard normal distribution. When the CAR procedures are implemented with $X_j, j = 1, 2$, I discretize them in the following way:

$$\tilde{x} = \begin{cases} 1 & \text{if } x < z_{p_j} \\ 0 & \text{if } x \geq z_{p_j} \end{cases},$$

where z_{p_j} is the p_j-quantile of the standard normal distribution. However, the original

continuous covariates will be included in the data analysis. I can see that my method can control the type I error rate very well, and estimate the parameters very accurately.

In Table 3.3, I report results for Pocock and Simon's design (PS) and complete randomization when both X_1 and X_2 are binary covariates with success rates of p_1 and p_2, respectively. I found that my proposed method can control the type I error rate very well and estimate the parameters very accurately. In Table 3.4, I report results for Pocock and Simon's design and complete randomization when both X_1 and X_2 follow standard normal distribution. I use the same way as in Table 3.2 to implement CAR. I get similar conclusion as in Table 3.3.

In Table 3.5, I offer results about the covariate imbalance for the scenario of Table 3.1. I report the average and standard deviation of the following values out of 10,000 replications: overall difference in patient numbers between the two treatments (DIF_n), the differences of patient numbers between the two treatments in the four stratum (DIF_{gh} for $X_1 = g$ and $X_2 = h$, $g, h = 0, 1$). In Table 3.6, I report results about the covariate imbalance for the scenario of Table 3.2. In this Table, DIF_{gh} refers to the stratum-level treatment assignment difference corresponding to the discretized covariates. I can see that compared to complete randomization, the overall and stratum imbalance can be controlled much better by my proposed method.

In Table 3.7, I report results about the covariate imbalance for the scenario of Table 3.3. In addition to the overall and stratum level imbalance, I also reported the marginal imbalance: $DIF_{1\cdot}$ is the marginal imbalance for $X_1 = 1$, $DIF_{0\cdot}$ is the marginal imbalance for $X_1 = 0$, $DIF_{\cdot 1}$ is the marginal imbalance for $X_2 = 1$, $DIF_{\cdot 0}$ is the marginal imbalance for $X_2 = 0$. I found that Pocock and Simon's design will return better balance in all levels: overall, marginal and stratum. Compared to the stratum imbalance, Pocock and Simon's design can control the marginal and overall imbalance better. In Table 3.8, I report results about the covariate imbalance for the scenario of Table 3.4. As in Table

3.6, the stratum and marginal level imbalance corresponds to the discretized covariates. I got similar conclusion as in Table 3.7.

3.4 Conclusion

Covariate-adaptive randomization designs including stratified permuted block randomization (Zelen, 1974) and Pocock and Simon's design (1975) are the most popular randomization design in the 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 chapter, I 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.

This chapter opens a door to future research topics. First, I consider the scenario that all the covariates used in the randomization procedures are used in the data analysis. However, in practice, clinical trial practitioners often use part of these randomization covariates or even just t-test in the data analysis. The reasons include: (i) researchers cannot explain the practical meaning of certain covariates effects; (ii) a large number of covariates in the model will lead to theoretical difficulties; (iii) the justification of the model specification becomes more difficult if more covariates are included in the model. I will study these scenarios in next chapter. Second, in this dissertation, I use the idea of Cui et al. (1999) to solve the problem of type I error rate, and offer lots of insight for other approaches such as the Fisher's product combination test proposed by Bauer and Köhne

(1994) and the weighted inverse normal method proposed by Lehmacher and Wassmer (1999). Third, Zhang et al. (2007) 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. The study on sequential monitoring of clinical trials with CARA and SSR is lacking in the literature. I leave all these for future research.

Table 3.1: Performance of SPB and complete randomization under H_0 when both covariates are discrete

$(\mu_1, \mu_2, \beta_1, \beta_2, p_1, p_2)$	Design	α	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_T$
$(0.5, 0.5, 1, 1, 0.5, 0.5)$	SPB	0.049	0.999(0.077)	1.00(0.077)	0.000(0.080)
$(0.5, 0.5, 1, 1, 0.5, 0.5)$	CR	0.051	1.00(0.079)	1.00(0.078)	0.001(0.080)
$(0.5, 0.5, 1, 1, 0.4, 0.6)$	SPB	0.051	1.00(0.079)	1.00(0.081)	0.001(0.080)
$(0.5, 0.5, 1, 1, 0.4, 0.6)$	CR	0.048	1.00(0.079)	0.999(0.080)	-0.001(0.080)
$(2, 2, 1, 1, 0.5, 0.5)$	SPB	0.053	0.999(0.078)	1.00(0.078)	0.000(0.082)
$(2, 2, 1, 1, 0.5, 0.5)$	CR	0.054	1.00(0.078)	0.999(0.078)	-0.001(0.082)
$(2, 2, 1, 1, 0.4, 0.6)$	SPB	0.050	0.999(0.079)	1.00(0.079)	-0.001(0.080)
$(2, 2, 1, 1, 0.4, 0.6)$	CR	0.050	1.00(0.079)	0.999(0.078)	0.000(0.079)

Note: $\mu_1, \mu_2, \beta_1, \beta_2$ are unknown parameters; p_1, p_2 are the success rates of binary covariates; α is the proportion of the number of rejections of H_0 out of 10,000 replications and the intended value is 0.05; $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\beta}_T$ are the average and standard deviation of the values out of $10,000$ replications.

Table 3.2: Performance of SPB and complete randomization under H_0 when both covariates are continuous

$(\mu_1, \mu_2, \beta_1, \beta_2, p_1, p_2)$	Design	α	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_T$
$(0.5, 0.5, 1, 1, 0.5, 0.5)$	SPB	0.055	0.999(0.039)	1.00(0.039)	0.000(0.081)
$(0.5, 0.5, 1, 1, 0.5, 0.5)$	CR	0.055	0.999(0.039)	1.00(0.039)	0.000(0.082)
$(0.5, 0.5, 1, 1, 0.4, 0.6)$	SPB	0.051	1.00(0.039)	1.00(0.039)	0.001(0.081)
$(0.5, 0.5, 1, 1, 0.4, 0.6)$	CR	0.052	0.999(0.039)	1.00(0.039)	0.000(0.080)
$(2, 2, 1, 1, 0.5, 0.5)$	SPB	0.053	1.00(0.039)	1.00(0.039)	0.000(0.081)
$(2, 2, 1, 1, 0.5, 0.5)$	CR	0.051	1.00(0.039)	1.00(0.039)	0.002(0.082)
$(2, 2, 1, 1, 0.4, 0.6)$	SPB	0.054	0.999(0.039)	1.00(0.039)	0.000(0.080)
$(2, 2, 1, 1, 0.4, 0.6)$	CR	0.057	1.00(0.039)	1.00(0.039)	0.000(0.082)

Table 3.3: Performance of Pocock and Simon's design and complete randomization under H_0 when both covariates are discrete

$(\mu_1, \mu_2, \beta_1, \beta_2, p_1, p_2)$	Design	α	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_T$
$(0.5, 0.5, 1, 1, 0.5, 0.5)$	PS	0.051	1.00(0.078)	1.00(0.078)	0.000(0.080)
$(0.5, 0.5, 1, 1, 0.5, 0.5)$	CR	0.051	1.00(0.079)	1.00(0.078)	0.001(0.080)
$(0.5, 0.5, 1, 1, 0.4, 0.6)$	PS	0.048	1.00(0.079)	0.999(0.080)	0.000(0.080)
$(0.5, 0.5, 1, 1, 0.4, 0.6)$	CR	0.048	1.00(0.079)	0.999(0.080)	-0.001(0.080)
$(2, 2, 1, 1, 0.5, 0.5)$	PS	0.051	1.00(0.077)	0.999(0.078)	-0.001(0.081)
$(2, 2, 1, 1, 0.5, 0.5)$	CR	0.054	1.00(0.078)	0.999(0.078)	-0.001(0.082)
$(2, 2, 1, 1, 0.4, 0.6)$	PS	0.051	1.00(0.079)	1.00(0.078)	0.000(0.081)
$(2, 2, 1, 1, 0.4, 0.6)$	CR	0.050	1.00(0.079)	0.999(0.078)	0.000(0.079)

Table 3.4: Performance of Pocock and Simon's design and complete randomization under H_0 when both covariates are continuous

$(\mu_1, \mu_2, \beta_1, \beta_2, p_1, p_2)$	Design	α	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_T$
$(0.5, 0.5, 1, 1, 0.5, 0.5)$	PS	0.049	0.999(0.039)	1.00(0.039)	-0.001(0.080)
$(0.5, 0.5, 1, 1, 0.5, 0.5)$	CR	0.055	0.999(0.039)	1.00(0.039)	0.000(0.082)
$(0.5, 0.5, 1, 1, 0.4, 0.6)$	PS	0.050	1.00(0.039)	1.00(0.039)	0.001(0.082)
$(0.5, 0.5, 1, 1, 0.4, 0.6)$	CR	0.052	0.999(0.039)	1.00(0.039)	0.000(0.080)
$(2, 2, 1, 1, 0.5, 0.5)$	PS	0.047	1.00(0.039)	1.00(0.038)	0.000(0.080)
$(2, 2, 1, 1, 0.5, 0.5)$	CR	0.051	1.00(0.039)	1.00(0.039)	0.002(0.082)
$(2, 2, 1, 1, 0.4, 0.6)$	PS	0.052	1.00(0.039)	1.00(0.039)	0.001(0.081)
$(2, 2, 1, 1, 0.4, 0.6)$	CR	0.057	1.00(0.039)	1.00(0.039)	0.000(0.082)

Table 3.5: Covariate imbalance for SPB and complete randomization when both covariates are discrete

$(\mu_1, \mu_2, \beta_1, \beta_2, p_1, p_2)$	Design	DIF_n	DIF_{11}	DIF_{10}	DIF_{01}	DIF_{00}
$(0.5, 0.5, 1, 1, 0.5, 0.5)$	SPB	1.32(1.27)	0.67(0.619)	0.664(0.626)	0.670(0.624)	0.672(0.622)
$(0.5, 0.5, 1, 1, 0.5, 0.5)$	CR	20.7(15.7)	10.4(7.91)	10.5(7.89)	10.6(8.11)	10.5(8.02)
$(0.5, 0.5, 1, 1, 0.4, 0.6)$	SPB	1.32(1.28)	0.660(0.621)	0.666(0.626)	0.669(0.618)	0.659(0.627)
$(0.5, 0.5, 1, 1, 0.4, 0.6)$	CR	20.8(16.1)	10.2(7.81)	8.35(6.44)	12.4(9.45)	10.2(7.87)
$(2, 2, 1, 1, 0.5, 0.5)$	SPB	1.32(1.28)	0.672(0.627)	0.660(0.623)	0.668(0.620)	0.663(0.619)
$(2, 2, 1, 1, 0.5, 0.5)$	CR	20.9(16.1)	10.4(7.98)	10.6(8.11)	10.4(7.89)	10.5(8.01)
$(2, 2, 1, 1, 0.4, 0.6)$	SPB	1.33(1.27)	0.662(0.620)	0.679(0.626)	0.674(0.624)	0.661(0.622)
$(2, 2, 1, 1, 0.4, 0.6)$	CR	20.8(16.2)	10.3(7.84)	8.42(6.40)	12.6(9.54)	10.3(7.83)

Note: The average and standard deviation of the following values out of 10,000 replications: overall difference in patient numbers between the two treatments (DIF_n), the differences of patient numbers between the two treatments in the four stratum (DIF_{gh} for $X_1 = g$ and $X_2 = h$, $g, h = 0, 1$).

Table 3.6: Covariate imbalance for SPB and complete randomization when both covariates are continuous

$(\mu_1, \mu_2, \beta_1, \beta_2, p_1, p_2)$	Design	DIF_n	DIF_{11}	DIF_{10}	DIF_{01}	DIF_{00}
$(0.5, 0.5, 1, 1, 0.5, 0.5)$	SPB	1.32(1.28)	0.678(0.623)	0.672(0.620)	0.660(0.622)	0.665(0.627)
$(0.5, 0.5, 1, 1, 0.5, 0.5)$	CR	20.8(16.0)	10.5(8.04)	10.5(8.01)	10.4(7.94)	10.4(7.99)
$(0.5, 0.5, 1, 1, 0.4, 0.6)$	SPB	1.32(1.27)	0.673(0.628)	0.673(0.623)	0.674(0.623)	0.669(0.631)
$(0.5, 0.5, 1, 1, 0.4, 0.6)$	CR	21.0(16.4)	10.4(7.87)	8.43(6.44)	12.5(9.69)	10.3(7.82)
$(2, 2, 1, 1, 0.5, 0.5)$	SPB	1.31(1.26)	0.671(0.624)	0.658(0.623)	0.660(0.622)	0.671(0.624)
$(2, 2, 1, 1, 0.5, 0.5)$	CR	20.9(16.0)	10.3(7.95)	10.5(8.06)	10.5(8.05)	10.4(7.96)
$(2, 2, 1, 1, 0.4, 0.6)$	SPB	1.33(1.27)	0.658(0.620)	0.672(0.629)	0.667(0.624)	0.668(0.620)
$(2, 2, 1, 1, 0.4, 0.6)$	CR	21.2(16.0)	10.0(7.79)	8.42(6.34)	12.7(9.67)	10.2(7.84)

Table 3.7: Covariate imbalance for Pocock and Simon's design and complete randomization when both covariates are discrete

$(\mu_1, \mu_2, \beta_1, \beta_2, p_1, p_2)$	Design	DIF_n	DIF_{11}	DIF_{10}	DIF_{01}	DIF_{00}	$DIF_{1\cdot}$	$DIF_{0\cdot}$	$DIF_{\cdot 1}$	$DIF_{\cdot 0}$
$(0.5, 0.5, 1, 1, 0.5, 0.5)$	PS	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)
$(0.5, 0.5, 1, 1, 0.5, 0.5)$	CR	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)
$(0.5, 0.5, 1, 1, 0.4, 0.6)$	PS	1.72(1.70)	5.17(3.92)	5.11(3.90)	5.18(3.96)	5.14(3.93)	1.55(1.39)	1.55(1.38)	1.58(1.41)	1.53(1.40)
$(0.5, 0.5, 1, 1, 0.4, 0.6)$	CR	20.8(16.1)	10.2(7.81)	8.35(6.44)	12.4(9.45)	10.2(7.87)	13.2(10.2)	16.2(12.3)	16.1(12.4)	13.1(10.2)
$(2, 2, 1, 1, 0.5, 0.5)$	PS	1.69(1.70)	5.43(4.16)	5.44(4.19)	5.42(4.14)	5.42(4.16)	1.55(1.42)	1.52(1.38)	1.53(1.38)	1.53(1.39)
$(2, 2, 1, 1, 0.5, 0.5)$	CR	20.9(16.1)	10.4(7.98)	10.6(8.11)	10.4(7.89)	10.5(8.01)	14.9(11.4)	14.7(11.3)	14.8(11.3)	14.9(11.3)
$(2, 2, 1, 1, 0.4, 0.6)$	PS	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)
$(2, 2, 1, 1, 0.4, 0.6)$	CR	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)

Note: The average and standard deviation of the following values out of 10,000 replications: overall difference in patient numbers between the two treatments (DIF_n); the differences of patient numbers between the two treatments in the four stratum (DIF_{gh} for $X_1 = g$ and $X_2 = h$, $g, h = 0, 1$); the marginal imbalance: $DIF_{1\cdot}$ is the marginal imbalance for $X_1 = 1$, $DIF_{0\cdot}$ is the marginal imbalance for $X_1 = 0$, $DIF_{\cdot 1}$ is the marginal imbalance for $X_2 = 1$, $DIF_{\cdot 0}$ is the marginal imbalance for $X_2 = 0$.

Table 3.8: Covariate imbalance for Pocock and Simon's design and complete randomization when both covariates are continuous

$(\mu_1, \mu_2, \beta_1, \beta_2, p_1, p_2)$	Design	DIF_n	DIF_{11}	DIF_{10}	DIF_{01}	DIF_{00}	$DIF_{1\cdot}$	$DIF_{0\cdot}$	$DIF_{\cdot 1}$	$DIF_{\cdot 0}$
$(0.5, 0.5, 1, 1, 0.5, 0.5)$	PS	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)
$(0.5, 0.5, 1, 1, 0.5, 0.5)$	CR	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)
$(0.5, 0.5, 1, 1, 0.4, 0.6)$	PS	1.69(1.66)	5.21(4.01)	5.12(3.94)	5.21(4.02)	5.15(3.99)	1.55(1.37)	1.54(1.37)	1.53(1.39)	1.54(1.38)
$(0.5, 0.5, 1, 1, 0.4, 0.6)$	CR	21.0(16.4)	10.4(7.87)	8.43(6.44)	12.5(9.69)	10.3(7.82)	13.3(10.2)	16.2(12.4)	16.4(12.4)	13.3(10.1)
$(2, 2, 1, 1, 0.5, 0.5)$	PS	1.69(1.69)	5.47(4.19)	5.44(4.17)	5.45(4.18)	5.43(4.17)	1.54(1.39)	1.53(1.37)	1.53(1.39)	1.54(1.38)
$(2, 2, 1, 1, 0.5, 0.5)$	CR	20.9(16.0)	10.3(7.95)	10.5(8.06)	10.5(8.05)	10.4(7.96)	14.8(11.4)	14.8(11.3)	14.8(11.3)	14.8(11.3)
$(2, 2, 1, 1, 0.4, 0.6)$	PS	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)
$(2, 2, 1, 1, 0.4, 0.6)$	CR	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)

Chapter 4

Sequential monitoring of randomized clinical trials with CAR and SSR-A subset of the randomization covariates are included in the data analysis

Abstract: In Chapter 3, I studied the sequential monitoring of covariate-adaptive randomized clinical trials with sample size re-estimation under the scenario where all the randomization covariates are included in the data analysis. That is recommended practice in clinical trials, but the comprehensive theoretical support is lacking in the literature. Therefore it is worth studying it and offering practical guidance for clinical trials. Another related but different topic is how to control the type I error rate when sequentially monitoring the covariate-adaptive randomized clinical trials with sample size re-estimation under the scenario where only a subset of the randomization covariates are included in

the data analysis. Numerical studies showed that the type I error rate is conservative, but in practice, clinical trial practitioners often do not include all the randomization covariates into the data analysis, which raised lots of concerns. Therefore, it is necessary to theoretically and numerically study this scenario. In this chapter, I proposed approaches to control the type I error rate, and performed theoretical and numerical studies on this procedure.

4.1 Introduction

The significance of covariate-adaptive randomization, sequential monitoring and sample size re-estimation have been introduced in Chapter 3. In this chapter, I discussed a situation raising lots of concerns. Theoretical and applied researchers all realized a common situation in real clinical trials: only some of the randomization covariates are included in the data analysis such as t-test. For example, Lai et al. (2006) studied the impact of music on maternal anxiety in kangaroos in a clinical trial where permuted block randomization stratified on gender was used to allocate the kangaroos and a t-test was used to perform the data analysis. There are many practical reasons for this scenario, (i) researchers cannot explain the practical meaning of certain covariates effects; (ii) a large number of covariates in the model will lead to theoretical difficulties; (iii) the justification of the model specification becomes more difficult if more covariates are included in the model.

Shao et al. (2010) is one of the most influential papers in this research topic, and they provided the following propositions: (1) a test that is valid under any fixed treatment allocation is valid under simple randomization and Efron's biased coin design; (2) analysis of covariance is valid if the covariates used in randomization are a function of the covariates used in the analysis. For linear regression with univariate covariate, they also proved that (3) the two-sample t-test under stratified randomization with Efron's biased

coin design employed within each stratum has a conservative type I error rate. Their explanation is that the stratified randomization procedure leads to dependence between the two samples and the variance estimator in the t-statistic ignores this correlation and overestimates the true variance of the estimator of the treatment effects. In addition, they proposed the bootstrap method to find an unbiased estimator for the true variance and the bootstrap t-test to control the correct type I error rate. Shao and Yu (2013) studied this topic for generalized linear models. Further, Ma et al. (2015) further generalized the above results to a family of CAR design and allow more covariates in the model.

Another influential paper in this field is Ma et al. (2015). Shao et al. (2010) has several limitations. First, they focused one special randomization design that does not include many other popular CAR designs such as minimization designs as special cases. Second, they focused on the linear model with only one covariate, which is obviously not enough in practice. Ma et al. (2015). addressed these two problems, and offered theoretical results for a general family of linear models with multiple covariates and a general family of CAR designs including the popular stratified permuted block randomization and the Pocock and Simon's design (1975). Their results are based on an easily satisfied condition that the difference in the patient numbers in the two treatment arms on any covariate margin is bounded in probability. This chapter will follow the framework of Ma et al. (2015).

In this chapter, I proposed a general framework for sequential monitoring clinical trials using CAR design for randomization, linear regression models with part of the randomization covariates or none of the randomization covariates (t-test) for analysis and SSR procedure. By simulation, I found that originally worked method in Chapter 3 will not work in this scenario and the type I error rate is conservative. Then I proposed numerical methods to fix this problem and control the type I error rate. I performed

extensive numerical studies and the results demonstrated that my proposed approaches can control the type I error rate very well.

In Section 4.2, I introduce the notations, framework, my proposed methods, and theoretical findings. In Section 4.3, I offer results from numerical results. Conclusions are in Section 4.4.

4.2 Sequential monitoring of CAR with SSR when part or none of the randomization covariates are in the data analysis

4.2.1 Framework

As in Chapter 3, assume that n originally planned subjects are sequentially allocated to a two-arm randomized controlled clinical trial 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). In this chapter, in addition to the covariates, $(X_1, \ldots, X_p)$, I introduce another sets of covariates, $(V_1, \ldots, V_q)$ to fit the scenario where part of the randomization covariates are omitted from the data analysis. That is, $(X_1, \ldots, X_p)$ represent the covariates used for both CAR design and data analysis, and $(V_1, \ldots, V_q)$ represent those covariates that are used for CAR, but are excluded for data analysis. Assume that all the covariates are independent and their expectations are all 0 without loss of generality, i.e., $E(X_{ik}) = 0, E(V_{ij}) = 0, i = 1, \ldots, n, k = 1, \ldots, p, j = 1, \ldots, q$. In addition, the errors are assumed to be independent with the covariates. Assume that the ith subject's response Y_i follows the linear model:

$$Y_i = \mu_1 T_i + \mu_2(1 - T_i) + X_{i1}\beta_1 + \ldots + X_{ip}\beta_p + V_{i1}\gamma_1 + \ldots + V_{iq}\gamma_q + \epsilon_i, \qquad (4.1)$$

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, I do not have to assume the errors follow normal distribution. I write $\boldsymbol{\mu} = (\mu_1, \mu_2)^T$, $\boldsymbol{\eta} = (\mu_1, \mu_2, \beta_1, \ldots, \beta_p)^T$, $\boldsymbol{T}(n) = (T_1, \ldots, T_n)^T$, $\boldsymbol{Y}(n) = (Y_1, \ldots, Y_n)^T$, $\boldsymbol{\epsilon}(n) = (\epsilon_1, \ldots, \epsilon_n)^T$ 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, I first discretize these continuous covariates, and apply CAR designs with respect to the discretized covariates. Specifically, let

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

and

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

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.

In this chapter, I perform the following hypothesis testing to compare two treatments in clinical trials:

$$H_0 : \mu_1 = \mu_2 \text{ versus } \mu_1 \neq \mu_2. \tag{4.2}$$

Let $\lfloor \cdot \rfloor$ denote the floor function and $t = N/n$ be the *information time* when N is the number of enrolled patients. There are two special cases when not all the randomization covariates are used in the data analysis. First, only part of the covariates, $(X_1, \ldots, X_p)$, are included in the data analysis. Then the sequential statistic to test the hypothesis (4.2) at time point $t \in (0, 1]$ is

$$Z_t = \frac{L\hat{\boldsymbol{\eta}}(t)}{\sqrt{\hat{\sigma}(t)^2 L(\boldsymbol{X}(\lfloor nt \rfloor)^T \boldsymbol{X}(\lfloor nt \rfloor))^{-1} L^T}}, \tag{4.3}$$

where $L = (1, -1, 0, \ldots, 0)$, $\hat{\boldsymbol{\eta}}(t) = (\boldsymbol{X}(\lfloor nt \rfloor)^T \boldsymbol{X}(\lfloor nt \rfloor))^{-1} \boldsymbol{X}(\lfloor nt \rfloor)^T \boldsymbol{Y}(\lfloor nt \rfloor)$, $\hat{\sigma}(t)^2 = [\boldsymbol{Y}(\lfloor nt \rfloor) - \boldsymbol{X}(\lfloor nt \rfloor)\hat{\boldsymbol{\eta}}(t)]^T [\boldsymbol{Y}(\lfloor nt \rfloor) - \boldsymbol{X}(\lfloor nt \rfloor)\hat{\boldsymbol{\eta}}(t)]/(\lfloor nt \rfloor - p - 2)$. Second, none of the covariates are used in the data analysis, which is the t-test or equivalently fitting the following model:

$$Y_i = \mu_1 T_i + \mu_2(1 - T_i) + \epsilon_i, i = 1, \ldots, n. \tag{4.4}$$

In this case, I do not have the covariates $(X_1, \ldots, X_p)$, and the responses follow:

$$Y_i = \mu_1 T_i + \mu_2(1 - T_i) + V_{i1}\gamma_1 + \ldots + V_{iq}\gamma_q + \epsilon_i, i = 1, \ldots, n. \tag{4.5}$$

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

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

Then the sequential statistic to test the hypothesis (4.2) at time point $t \in (0, 1]$ is

$$Z'_t = \frac{Q\hat{\boldsymbol{\mu}}(t)}{\sqrt{\hat{\sigma}(t)^2 Q(\boldsymbol{Tr}(\lfloor nt \rfloor)^T \boldsymbol{Tr}(\lfloor nt \rfloor))^{-1} Q^T}}, \tag{4.6}$$

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

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

4.2.2 Incorporation of sample size re-estimation

In this chapter, I use the same SSR approach as in Chapter 3. The following sequential statistics were used in Chapter 3 and the type I error rate was successfully controlled when all the randomization covariates are included in the data analysis.

$$U_t = \begin{cases} Z_t, & \text{if } t \le t_L; \\ [w(t_L, t)]^{1/2} \times Z_{t_L} + [1 - w(t_L, t)]^{1/2} \times & \\ \quad \{[B(b(t - t_L) + t_L) - B(t_L)]/[b(t - t_L)]^{1/2}\}, & \text{if } t > t_L, \end{cases} \tag{4.7}$$

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

We first perform numerical study to investigate whether the same method can work when not all the randomization covariates are included in the data analysis. For all the tables, suppose originally planned 500 patients sequentially enter a clinical trial, and the responses follow

$$Y_i = \mu_1 T_i + \mu_2(1 - T_i) + X_{i1}\beta_1 + X_{i2}\beta_2 + \epsilon_i, i = 1, \ldots, 500, \tag{4.8}$$

where $(\mu_1, \mu_2, \beta_1, \beta_2)$ are unknown parameters, and ϵ_i are independent errors from the normal distribution $N(0, 1)$. Here I do not distinguish the notation X and V to save space. The CAR designs will be applied with respect to both X_1 and X_2, and different distributions of these two covariates will be considered. I 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. In this case, $w(t_2, t_3) = 0.5$ and $b_{\max} = 2$. All the results are based on 10,000 replications.

In Table 4.1, I report results for SPB when both X_1 and X_2 are binary covariates with success rates of p_1 and p_2, respectively, and only X_1 is included in the working model as follows

$$Y_i = \beta_0 + \beta_T T_i + X_{i1}\beta_1 + \epsilon_i, i = 1, \ldots, 500. \tag{4.9}$$

I offer results for type I error rate (α), average and standard deviation of estimates of β_1 and β_T out of 10,000 replications. In Table 4.2, I report results for SPB when both X_1 and X_2 follow standard normal distribution and only X_1 is included in the working model as in Table 1. When the CAR procedures are implemented with $X_j, j = 1, 2$, I discretize them in the following way:

$$\tilde{x} = \begin{cases} 1 & \text{if } x < z_{p_j} \\ 0 & \text{if } x \geq z_{p_j} \end{cases},$$

where z_{p_j} is the p_j-quantile of the standard normal distribution. However, the original continuous covariates will be included in the data analysis. I get similar conclusions as in Table 4.1. In Table 4.3, I report results for Pocock and Simon's design (PS) when both X_1 and X_2 are binary covariates with success rates of p_1 and p_2, respectively and only X_1 is included in the working model as in Table 4.1. In Table 4.4, I report results for Pocock and Simon's design when both X_1 and X_2 follow standard normal distribution and only X_1 is included in the working model as in Table 4.1. I use the same way as in Table 4.2 to implement CAR. In all the tables, I can see that the type I error rates are all conservative. But I can still estimate the parameters very accurately. In Tables 4.5-4.8, I perform numerical study for the similar scenarios to Tables 4.1-4.4, but use

t-test statistics in the data analysis. I found the type I error rates are more conservative but the unknown parameters can be estimated very well.

4.2.3 Asymptotic results

In this chapter, I propose to revise Z_t and Z_t' and the corresponding U_t to control the type I error rate.

Let

$$Z_t^{adj} = \frac{L\hat{\boldsymbol{\eta}}(t)}{\hat{\epsilon}(t)\sqrt{\hat{\sigma}(t)^2 L(\boldsymbol{X}(\lfloor nt \rfloor)^T \boldsymbol{X}(\lfloor nt \rfloor))^{-1} L^T}}, \tag{4.10}$$

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

$$\frac{\sum\limits_{j \in C^*} \gamma_j^2 \sigma_{\delta j}^2 + \sigma^2}{\sigma^2 + \sum\limits_{j=1}^{q} Var(V_j \gamma_j^T)}, \tag{4.11}$$

$\sigma_{\delta j}^2 = E\left[Var\left(\delta_j | d_j^*(V_j)\right)\right]$, and $\delta_j = V_j - E(V_j | d_j^*(V_j))$. Then I have the following theorem for the scenario when part of the randomization covariates are included in the data analysis.

Theorem 4.1. *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. The sequence of test statistics $\{(U_{t_1}, \ldots, U_{t_K}), 0 \le t_1 \le t_2 \le \ldots \le t_K \le 1\}$ has the asymptotic canonical joint distribution defined in Jennison and Turnbull (2000), i.e., under H_0,*

(i) $\{U_{t_1}, \ldots, U_{t_K}\}$ follows multivariate normal distribution;

(ii) $EU_{t_i} = 0$;

(iii) $Cov(U_{t_i}, U_{t_j}) = \sqrt{\lfloor nt_i \rfloor / \lfloor nt_j \rfloor}$, $0 \le t_i \le t_j \le 1$.

When t-test is used in the data analysis, I revise Z_t' in the following way and the

corresponding U_t can be calculated. Let

$$Z_t^{adj'} = \frac{Q\hat{\boldsymbol{\mu}}(t)}{\hat{\epsilon}(t)\sqrt{\hat{\sigma}(t)^2 Q(\boldsymbol{Tr}(\lfloor nt \rfloor)^T \boldsymbol{Tr}(\lfloor nt \rfloor))^{-1} Q^T}}, \tag{4.12}$$

where $\hat{\boldsymbol{\mu}}(t) = (\boldsymbol{Tr}(\lfloor nt \rfloor)^T \boldsymbol{Tr}(\lfloor nt \rfloor))^{-1} \boldsymbol{Tr}(\lfloor nt \rfloor)^T \boldsymbol{Y}(\lfloor nt \rfloor)$,
$\hat{\sigma}(t)^2 = [\boldsymbol{Y}(\lfloor nt \rfloor) - \boldsymbol{Tr}(\lfloor nt \rfloor)\hat{\boldsymbol{\mu}}(t)]^T [\boldsymbol{Y}(\lfloor nt \rfloor) - \boldsymbol{Tr}(\lfloor nt \rfloor)\hat{\boldsymbol{\mu}}(t)]/(\lfloor nt \rfloor - 2)$, and $\hat{\epsilon}(t)^2$ 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 I have the following theorem.

Theorem 4.2. *Let $B_t^U = \sqrt{t} U_t$. 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^U is asymptotically a standard Brownian motion in distribution. The sequence of test statistics $\{(U_{t_1}, \ldots, U_{t_K}), 0 \le t_1 \le t_2 \le \ldots \le t_K \le 1\}$ has the asymptotic canonical joint distribution defined in Jennison and Turnbull (2000), i.e., under H_0,*

(i) $\{U_{t_1}, \ldots, U_{t_K}\}$ follows multivariate normal distribution;

(ii) $EU_{t_i} = 0$;

(iii) $Cov(U_{t_i}, U_{t_j}) = \sqrt{\lfloor nt_i \rfloor / \lfloor nt_j \rfloor}$, $0 \le t_i \le t_j \le 1$.

This theorem reveals the most fundamental properties for the proposed method, i.e., the asymptotic joint distribution of the sequential statistics. From this theorem and the numerical studies above, I can easily see and understand the conservativeness of the type I error rates when not all the randomization covariates are included in the data analysis, 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.

4.3 Numerical and simulation studies

In this section, I study the finite-sample properties of the proposed procedure. For all the tables, suppose originally planned 500 patients sequentially enter a clinical trial, and the responses follow

$$Y_i = \mu_1 T_i + \mu_2(1 - T_i) + X_{i1}\beta_1 + X_{i2}\beta_2 + \epsilon_i, i = 1, \ldots, 500, \tag{4.13}$$

where $(\mu_1, \mu_2, \beta_1, \beta_2)$ are unknown parameters, and ϵ_i are independent errors from the normal distribution $N(0, 1)$. In different tables, I compare the stratified permuted block randomization (SPB), Pocock and Simon's procedure (PS) and complete randomization. The CAR designs will be applied with respect to both X_1 and X_2, and different distributions of these two covariates will be considered. I 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. In this case, $w(t_2, t_3) = 0.5$ and $b_{\max} = 2$. It is worth noting that a variety of approaches such as bootstraps can be used to obtain $\hat{\epsilon}$. In this dissertation, I obtain $\hat{\epsilon}$ in the following way. At each interim look, I fit model (4.13) with full data to obtain consistent estimators of the unknown parameters. I 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 (Lehmann, 2004). All the results are based on 10,000 replications.

In Tables 4.9-4.12, the sequential data analysis are all based on the model (4.9), and the adjusted sequential statistics U_t are used. In Table 4.9, I report results for SPB and complete randomization when both X_1 and X_2 are binary covariates with success rates of p_1 and p_2, respectively. I offer results for type I error rate (α) (the proportion of the number of rejections of H_0 out of 10,000 replications, and the intended value is 0.05), average and standard deviation of the following values out of 10,000 replications:

estimates of β_1 and β_T. I can see that my method can control the type I error rate very well, and estimate the parameters very accurately.

In Table 4.10, I report results for SPB and complete randomization when both X_1 and X_2 follow standard normal distribution. When the CAR procedures are implemented with $X_j, j = 1, 2$, I discretize them in the following way:

$$\tilde{x} = \begin{cases} 1 & \text{if } x < z_{p_j} \\ 0 & \text{if } x \geq z_{p_j} \end{cases},$$

where z_{p_j} is the p_j-quantile of the standard normal distribution. However, the original continuous covariates will be included in the data analysis. I get similar conclusions as in Table 4.9.

In Table 4.11, I report results for Pocock and Simon's design (PS) and complete randomization when both X_1 and X_2 are binary covariates with success rates of p_1 and p_2, respectively. I found that my proposed method can control the type I error rate very well and estimate the parameters very accurately. In Table 4.12, I report results for Pocock and Simon's design and complete randomization when both X_1 and X_2 follow standard normal distribution. I use the same way as in Table 4.10 to implement CAR. I get similar conclusion as in Table 4.11.

In Table 4.13, I report the covariate imbalance for the scenario of Table 4.9. I report the average and standard deviation of the following values out of 10,000 replications: overall difference in patient numbers between the two treatments (DIF_n), the differences of patient numbers between the two treatments in the four stratum (DIF_{gh} for $X_1 = g$ and $X_2 = h$, $g, h = 0, 1$). In Table 4.14, I report the covariate imbalance for the scenario of Table 4.10. In this table, DIF_{gh} refers to the stratum-level treatment assignment difference corresponding to the discretized covariates. In these two tables, I can see that compared to complete randomization, the overall and stratum imbalance can

be controlled much better by my proposed method. In Table 4.15, I report the covariate imbalance for the scenario of Table 4.11. I additionally report the marginal imbalance: $DIF_{1\cdot}$ is the marginal imbalance for $X_1 = 1$, $DIF_{0\cdot}$ is the marginal imbalance for $X_1 = 0$, $DIF_{\cdot 1}$ is the marginal imbalance for $X_2 = 1$, $DIF_{\cdot 0}$ is the marginal imbalance for $X_2 = 0$. I found that Pocock and Simon's design will return better balance in all levels: overall, marginal and stratum. Compared to the stratum imbalance, Pocock and Simon's design can control the marginal and overall imbalance better. In Table 4.16, I report the covariate imbalance for the scenario of Table 4.12. In this table, the stratum and marginal level imbalance corresponds to the discretized covariates. I got similar results to Table 4.15.

In Tables 4.17-4.20, I report results when the sequential data analysis are all based on the t-test and the adjusted sequential statistics U_t are used. In these tables, I get similar corresponding results as when I include only one covariate in the model for data analysis.

4.4 Conclusion

The advantages and challenges of the combination of covariate-adaptive randomization, sequential monitoring and sample size re-estimation have been introduced in Chapter 3. In practice, clinical trial practitioners are often reluctant to include all the randomization covariates in the data analysis for different reasons, which introduces extra problems. In this chapter, I study the theoretical and numerical properties for sequential monitoring of covariate-adaptive clinical trials with sample size re-estimation when not all the randomization covariates are included in the data analysis. I found that using the approaches in Chapter 3 without adjustment will lead to conservative type I error rate. The lower number of randomization covariates I include in the data analysis, the more conservative the type error rate is. I proposed methods to adjust the sequential

test statistics based on the theoretical results, and successfully controlled the type I error rate demonstrated by the numerical study.

In addition to the future research fields mentioned in the conclusion of Chapter 3, 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. Other methods leading to a correct estimation of this variance can also be investigated. The same problem occurs when CAR is used and longitudinal data analysis is implemented. I leave all these for future research.

Table 4.1: Performance for SPB under H_0 when both covariates are discrete and only one covariate is included in the data analysis

$(\mu_1, \mu_2, \beta_1, \beta_2, p_1, p_2)$	Design	α	$\hat{\beta}_1$	$\hat{\beta}_T$
$(0.5, 0.5, 1, 1, 0.5, 0.5)$	SPB	0.027	1.00(0.086)	0.000(0.078)
$(0.5, 0.5, 1, 1, 0.4, 0.6)$	SPB	0.029	0.999(0.088)	0.001(0.078)
$(2, 2, 1, 1, 0.5, 0.5)$	SPB	0.032	0.999(0.088)	0.000(0.079)
$(2, 2, 1, 1, 0.4, 0.6)$	SPB	0.030	0.999(0.087)	-0.001(0.077)

Table 4.2: Performance for SPB under H_0 when both covariates are continuous and only one covariate is included in the data analysis

$(\mu_1, \mu_2, \beta_1, \beta_2, p_1, p_2)$	Design	α	$\hat{\beta}_1$	$\hat{\beta}_T$
$(0.5, 0.5, 1, 1, 0.5, 0.5)$	SPB	0.019	0.999(0.055)	-0.002(0.089)
$(0.5, 0.5, 1, 1, 0.4, 0.6)$	SPB	0.017	1.00(0.054)	0.001(0.089)
$(2, 2, 1, 1, 0.5, 0.5)$	SPB	0.019	1.00(0.055)	0.000(0.091)
$(2, 2, 1, 1, 0.4, 0.6)$	SPB	0.018	0.999(0.055)	-0.001(0.090)

Table 4.3: Performance for Pocock and Simon's design under H_0 when both covariates are discrete and only one covariate is included in the data analysis

$(\mu_1, \mu_2, \beta_1, \beta_2, p_1, p_2)$	Design	α	$\hat{\beta}_1$	$\hat{\beta}_T$
$(0.5, 0.5, 1, 1, 0.5, 0.5)$	PS	0.030	1.00(0.087)	0.000(0.077)
$(0.5, 0.5, 1, 1, 0.4, 0.6)$	PS	0.028	1.00(0.087)	0.000(0.078)
$(2, 2, 1, 1, 0.5, 0.5)$	PS	0.028	1.00(0.087)	-0.001(0.078)
$(2, 2, 1, 1, 0.4, 0.6)$	PS	0.029	1.00(0.088)	0.000(0.078)

Table 4.4: Performance for Pocock and Simon's design under H_0 when both covariates are continuous and only one covariate is included in the data analysis

$(\mu_1, \mu_2, \beta_1, \beta_2, p_1, p_2)$	Design	α	$\hat{\beta}_1$	$\hat{\beta}_T$
$(0.5, 0.5, 1, 1, 0.5, 0.5)$	PS	0.019	0.999(0.055)	-0.001(0.090)
$(0.5, 0.5, 1, 1, 0.4, 0.6)$	PS	0.018	1.00(0.055)	0.001(0.091)
$(2, 2, 1, 1, 0.5, 0.5)$	PS	0.018	1.00(0.056)	-0.001(0.090)
$(2, 2, 1, 1, 0.4, 0.6)$	PS	0.023	1.00(0.054)	0.000(0.091)

Table 4.5: Performance for SPB under H_0 when both covariates are discrete and t-test is used

$(\mu_1, \mu_2, \beta_1, \beta_2, p_1, p_2)$	Design	α
$(0.5, 0.5, 1, 1, 0.5, 0.5)$	SPB	0.013
$(0.5, 0.5, 1, 1, 0.4, 0.6)$	SPB	0.013
$(2, 2, 1, 1, 0.5, 0.5)$	SPB	0.010
$(2, 2, 1, 1, 0.4, 0.6)$	SPB	0.011

Table 4.6: Performance for SPB under H_0 when both covariates are continuous and t-test is used

$(\mu_1, \mu_2, \beta_1, \beta_2, p_1, p_2)$	Design	α
$(0.5, 0.5, 1, 1, 0.5, 0.5)$	SPB	0.006
$(0.5, 0.5, 1, 1, 0.4, 0.6)$	SPB	0.008
$(2, 2, 1, 1, 0.5, 0.5)$	SPB	0.007
$(2, 2, 1, 1, 0.4, 0.6)$	SPB	0.006

Table 4.7: Performance for Pocock and Simon's design under H_0 when both covariates are discrete and t-test is used

$(\mu_1, \mu_2, \beta_1, \beta_2, p_1, p_2)$	Design	α
$(0.5, 0.5, 1, 1, 0.5, 0.5)$	PS	0.015
$(0.5, 0.5, 1, 1, 0.4, 0.6)$	PS	0.014
$(2, 2, 1, 1, 0.5, 0.5)$	PS	0.016
$(2, 2, 1, 1, 0.4, 0.6)$	PS	0.012

Table 4.8: Performance for Pocock and Simon's design under H_0 when both covariates are continuous and t-test is used

$(\mu_1, \mu_2, \beta_1, \beta_2, p_1, p_2)$	Design	α
$(0.5, 0.5, 1, 1, 0.5, 0.5)$	PS	0.007
$(0.5, 0.5, 1, 1, 0.4, 0.6)$	PS	0.010
$(2, 2, 1, 1, 0.5, 0.5)$	PS	0.006
$(2, 2, 1, 1, 0.4, 0.6)$	PS	0.007

Table 4.9: Performance for SPB and complete randomization under H_0 when both covariates are discrete and partial covariates are used. Adjusted statistics for SPB and unadjusted statistics for complete randomization are used for data analysis.

$(\mu_1, \mu_2, \beta_1, \beta_2, p_1, p_2)$	Design	α	$\hat{\beta}_1$	$\hat{\beta}_T$
$(0.5, 0.5, 1, 1, 0.5, 0.5)$	SPB	0.050	1.00(0.086)	0.000(0.080)
$(0.5, 0.5, 1, 1, 0.5, 0.5)$	CR	0.050	1.00(0.087)	0.001(0.089)
$(0.5, 0.5, 1, 1, 0.4, 0.6)$	SPB	0.052	1.00(0.088)	0.001(0.080)
$(0.5, 0.5, 1, 1, 0.4, 0.6)$	CR	0.049	1.00(0.087)	-0.001(0.088)
$(2, 2, 1, 1, 0.5, 0.5)$	SPB	0.053	1.00(0.088)	-0.001(0.082)
$(2, 2, 1, 1, 0.5, 0.5)$	CR	0.056	1.00(0.087)	-0.001(0.091)
$(2, 2, 1, 1, 0.4, 0.6)$	SPB	0.052	1.00(0.087)	-0.001(0.080)
$(2, 2, 1, 1, 0.4, 0.6)$	CR	0.052	1.00(0.088)	0.001(0.090)

Table 4.10: Performance for SPB and complete randomization under H_0 when both covariates are continuous and partial covariates are used. Adjusted statistics for SPB and unadjusted statistics for complete randomization are used for data analysis.

$(\mu_1, \mu_2, \beta_1, \beta_2, p_1, p_2)$	Design	α	$\hat{\beta}_1$	$\hat{\beta}_T$
$(0.5, 0.5, 1, 1, 0.5, 0.5)$	SPB	0.049	0.999(0.055)	-0.002(0.094)
$(0.5, 0.5, 1, 1, 0.5, 0.5)$	CR	0.052	0.999(0.055)	0.001(0.115)
$(0.5, 0.5, 1, 1, 0.4, 0.6)$	SPB	0.055	1.00(0.055)	0.002(0.095)
$(0.5, 0.5, 1, 1, 0.4, 0.6)$	CR	0.049	1.00(0.055)	0.000(0.113)
$(2, 2, 1, 1, 0.5, 0.5)$	SPB	0.056	1.00(0.055)	0.000(0.095)
$(2, 2, 1, 1, 0.5, 0.5)$	CR	0.053	1.00(0.055)	0.002(0.115)
$(2, 2, 1, 1, 0.4, 0.6)$	SPB	0.054	0.999(0.055)	-0.001(0.095)
$(2, 2, 1, 1, 0.4, 0.6)$	CR	0.051	1.00(0.055)	-0.001(0.115)

Table 4.11: Performance for Pocock and Simon's design and complete randomization when both covariates are discrete and partial covariates is used. Adjusted statistics for Pocock and Simon's design and unadjusted statistics for complete randomization are used for data analysis.

$(\mu_1, \mu_2, \beta_1, \beta_2, p_1, p_2)$	Design	α	$\hat{\beta}_1$	$\hat{\beta}_T$
$(0.5, 0.5, 1, 1, 0.5, 0.5)$	SPB	0.052	1.00(0.087)	0.000(0.081)
$(0.5, 0.5, 1, 1, 0.5, 0.5)$	CR	0.050	1.00(0.087)	0.001(0.089)
$(0.5, 0.5, 1, 1, 0.4, 0.6)$	SPB	0.050	1.00(0.088)	0.000(0.081)
$(0.5, 0.5, 1, 1, 0.4, 0.6)$	CR	0.049	1.00(0.087)	-0.001(0.088)
$(2, 2, 1, 1, 0.5, 0.5)$	SPB	0.052	1.00(0.087)	-0.001(0.081)
$(2, 2, 1, 1, 0.5, 0.5)$	CR	0.056	1.00(0.087)	-0.001(0.091)
$(2, 2, 1, 1, 0.4, 0.6)$	SPB	0.052	1.00(0.089)	0.000(0.081)
$(2, 2, 1, 1, 0.4, 0.6)$	CR	0.052	1.00(0.088)	0.001(0.090)

Table 4.12: Performance for Pocock and Simon's design and complete randomization when both covariates are continuous and partial covariates is used. Adjusted statistics for Pocock and Simon's design and unadjusted statistics for complete randomization are used for data analysis.

$(\mu_1, \mu_2, \beta_1, \beta_2, p_1, p_2)$	Design	α	$\hat{\beta}_1$	$\hat{\beta}_T$
$(0.5, 0.5, 1, 1, 0.5, 0.5)$	PS	0.051	0.999(0.055)	-0.001(0.095)
$(0.5, 0.5, 1, 1, 0.5, 0.5)$	CR	0.052	0.999(0.055)	0.001(0.115)
$(0.5, 0.5, 1, 1, 0.4, 0.6)$	PS	0.055	0.999(0.055)	0.001(0.096)
$(0.5, 0.5, 1, 1, 0.4, 0.6)$	CR	0.049	1.00(0.055)	0.000(0.113)
$(2, 2, 1, 1, 0.5, 0.5)$	PS	0.049	1.00(0.056)	-0.001(0.094)
$(2, 2, 1, 1, 0.5, 0.5)$	CR	0.053	1.00(0.055)	0.002(0.115)
$(2, 2, 1, 1, 0.4, 0.6)$	PS	0.058	1.00(0.055)	0.001(0.096)
$(2, 2, 1, 1, 0.4, 0.6)$	CR	0.051	1.00(0.055)	-0.001(0.115)

Table 4.13: Covariates imbalance for SPB and complete randomization when both covariates are discrete and partial covariates is used. Adjusted statistics for SPB and unadjusted statistics for complete randomization are used for data analysis.

$(\mu_1, \mu_2, \beta_1, \beta_2, p_1, p_2)$	Design	DIF_n	DIF_{11}	DIF_{10}	DIF_{01}	DIF_{00}
$(0.5, 0.5, 1, 1, 0.5, 0.5)$	SPB	1.32(1.27)	0.669(0.619)	0.666(0.627)	0.676(0.625)	0.669(0.622)
$(0.5, 0.5, 1, 1, 0.5, 0.5)$	CR	20.8(15.8)	10.4(7.89)	10.5(7.92)	10.6(8.10)	10.5(8.06)
$(0.5, 0.5, 1, 1, 0.4, 0.6)$	SPB	1.31(1.28)	0.658(0.622)	0.667(0.625)	0.667(0.617)	0.658(0.626)
$(0.5, 0.5, 1, 1, 0.4, 0.6)$	CR	20.9(16.0)	10.3(7.83)	8.38(6.41)	12.4(9.38)	10.2(7.86)
$(2, 2, 1, 1, 0.5, 0.5)$	SPB	1.32(1.28)	0.670(0.627)	0.661(0.622)	0.668(0.622)	0.665(0.617)
$(2, 2, 1, 1, 0.5, 0.5)$	CR	20.9(15.9)	10.4(7.95)	10.5(8.07)	10.4(7.92)	10.4(8.00)
$(2, 2, 1, 1, 0.4, 0.6)$	SPB	1.33(1.27)	0.664(0.621)	0.678(0.627)	0.678(0.627)	0.661(0.622)
$(2, 2, 1, 1, 0.4, 0.6)$	CR	20.9(16.1)	10.3(7.83)	8.38(6.35)	12.5(9.56)	10.2(7.83)

Table 4.14: Covariates imbalance for SPB and complete randomization when both covariates are continuous and partial covariates is used. Adjusted statistics for SPB and unadjusted statistics for complete randomization are used for data analysis.

$(\mu_1, \mu_2, \beta_1, \beta_2, p_1, p_2)$	Design	DIF_n	DIF_{11}	DIF_{10}	DIF_{01}	DIF_{00}
$(0.5, 0.5, 1, 1, 0.5, 0.5)$	SPB	1.31(1.27)	0.669(0.620)	0.665(0.620)	0.663(0.619)	0.667(0.630)
$(0.5, 0.5, 1, 1, 0.5, 0.5)$	CR	20.8(16.0)	10.5(8.04)	10.5(7.97)	10.4(7.91)	10.4(7.98)
$(0.5, 0.5, 1, 1, 0.4, 0.6)$	SPB	1.32(1.27)	0.673(0.628)	0.670(0.624)	0.672(0.622)	0.666(0.631)
$(0.5, 0.5, 1, 1, 0.4, 0.6)$	CR	20.9(16.3)	10.3(7.82)	8.41(6.39)	12.6(9.72)	10.3(7.80)
$(2, 2, 1, 1, 0.5, 0.5)$	SPB	1.31(1.25)	0.674(0.623)	0.659(0.621)	0.656(0.618)	0.665(0.621)
$(2, 2, 1, 1, 0.5, 0.5)$	CR	21.0(16.1)	10.4(7.97)	10.5(7.97)	10.6(8.02)	10.4(7.97)
$(2, 2, 1, 1, 0.4, 0.6)$	SPB	1.31(1.27)	0.660(0.623)	0.671(0.627)	0.671(0.624)	0.667(0.618)
$(2, 2, 1, 1, 0.4, 0.6)$	CR	21.2(16.1)	10.0(7.79)	8.44(6.38)	12.6(9.58)	10.3(7.85)

Table 4.15: Covariates imbalance for Pocock and Simon's design and complete randomization when both covariates are discrete and partial covariates is used. Adjusted statistics for Pocock and Simon's design and unadjusted statistics for complete randomization are used for data analysis.

$(\mu_1, \mu_2, \beta_1, \beta_2, p_1, p_2)$	Design	DIF_n	DIF_{11}	DIF_{10}	DIF_{01}	DIF_{00}	$DIF_{1\cdot}$	$DIF_{0\cdot}$	$DIF_{\cdot 1}$	$DIF_{\cdot 0}$
$(0.5, 0.5, 1, 1, 0.5, 0.5)$	PS	1.70(1.71)	5.44(4.16)	5.45(4.16)	5.46(4.17)	5.46(4.17)	1.54(1.42)	1.53(1.38)	1.54(1.38)	1.54(1.40)
$(0.5, 0.5, 1, 1, 0.5, 0.5)$	CR	20.8(15.8)	10.4(7.89)	10.5(7.92)	10.6(8.1)	10.5(8.06)	14.5(11.1)	14.9(11.5)	14.9(11.4)	14.8(11.2)
$(0.5, 0.5, 1, 1, 0.4, 0.6)$	PS	1.73(1.69)	5.17(3.93)	5.11(3.90)	5.19(3.97)	5.14(3.93)	1.54(1.39)	1.55(1.39)	1.58(1.42)	1.53(1.40)
$(0.5, 0.5, 1, 1, 0.4, 0.6)$	CR	20.9(16.0)	10.3(7.83)	8.38(6.41)	12.4(9.38)	10.2(7.86)	13.3(10.2)	16.2(12.3)	16.1(12.3)	13.2(10.2)
$(2, 2, 1, 1, 0.5, 0.5)$	PS	1.68(1.69)	5.43(4.16)	5.44(4.20)	5.44(4.14)	5.43(4.17)	1.54(1.42)	1.50(1.38)	1.53(1.38)	1.54(1.39)
$(2, 2, 1, 1, 0.5, 0.5)$	CR	20.9(15.9)	10.4(7.95)	10.5(8.07)	10.4(7.92)	10.4(8.00)	14.9(11.3)	14.7(11.3)	14.7(11.2)	14.9(11.4)
$(2, 2, 1, 1, 0.4, 0.6)$	PS	1.68(1.66)	5.21(4.00)	5.19(3.97)	5.25(4.04)	5.21(4.00)	1.52(1.40)	1.54(1.38)	1.51(1.37)	1.51(1.39)
$(2, 2, 1, 1, 0.4, 0.6)$	CR	20.9(16.1)	10.3(7.83)	8.38(6.35)	12.5(9.56)	10.2(7.83)	13.3(10.1)	16.3(12.4)	16.3(12.3)	13.2(10.1)

Table 4.16: Covariates imbalance for Pocock and Simon's design and complete randomization when both covariates are continuous and partial covariates is used. Adjusted statistics for Pocock and Simon's design and unadjusted statistics for complete randomization are used for data analysis.

$(\mu_1, \mu_2, \beta_1, \beta_2, p_1, p_2)$	Design	DIF_n	DIF_{11}	DIF_{10}	DIF_{01}	DIF_{00}	$DIF_{1\cdot}$	$DIF_{0\cdot}$	$DIF_{\cdot 1}$	$DIF_{\cdot 0}$
$(0.5, 0.5, 1, 1, 0.5, 0.5)$	PS	1.70(1.68)	5.42(4.18)	5.45(4.17)	5.41(4.14)	5.42(4.17)	1.54(1.40)	1.55(1.39)	1.53(1.40)	1.55(1.39)
$(0.5, 0.5, 1, 1, 0.5, 0.5)$	CR	20.8(16.0)	10.5(8.04)	10.5(7.97)	10.4(7.91)	10.4(7.98)	14.8(11.4)	14.7(11.3)	14.8(11.4)	14.8(11.3)
$(0.5, 0.5, 1, 1, 0.4, 0.6)$	PS	1.69(1.66)	5.21(4.01)	5.15(3.97)	5.24(4.03)	5.19(3.99)	1.53(1.39)	1.55(1.41)	1.55(1.40)	1.53(1.37)
$(0.5, 0.5, 1, 1, 0.4, 0.6)$	CR	20.9(16.3)	10.3(7.82)	8.41(6.39)	12.6(9.72)	10.3(7.80)	13.3(10.2)	16.2(12.4)	16.3(12.4)	13.3(10.1)
$(2, 2, 1, 1, 0.5, 0.5)$	PS	1.68(1.67)	5.43(4.22)	5.43(4.17)	5.40(4.20)	5.38(4.18)	1.52(1.38)	1.52(1.39)	1.51(1.37)	1.52(1.36)
$(2, 2, 1, 1, 0.5, 0.5)$	CR	21.0(16.1)	10.4(7.97)	10.5(7.97)	10.6(8.02)	10.4(7.97)	14.8(11.5)	14.8(11.3)	15(11.3)	14.9(11.3)
$(2, 2, 1, 1, 0.4, 0.6)$	PS	1.70(1.70)	5.22(3.96)	5.16(3.90)	5.24(3.97)	5.19(3.93)	1.54(1.42)	1.52(1.40)	1.53(1.39)	1.55(1.42)
$(2, 2, 1, 1, 0.4, 0.6)$	CR	21.2(16.1)	10.0(7.79)	8.44(6.38)	12.6(9.58)	10.3(7.85)	13.2(10.1)	16.4(12.6)	16.2(12.2)	13.3(10.2)

Table 4.17: Performance for SPB and complete randomization under H_0 when both covariates are discrete and t-test are used. Adjusted statistics for SPB and unadjusted statistics for complete randomization are used for data analysis.

$(\mu_1, \mu_2, \beta_1, \beta_2, p_1, p_2)$	Design	α	DIF_n	DIF_{11}	DIF_{10}	DIF_{01}	DIF_{00}
$(0.5, 0.5, 1, 1, 0.5, 0.5)$	SPB	0.052	1.32(1.27)	0.671(0.618)	0.666(0.627)	0.676(0.624)	0.668(0.621)
$(0.5, 0.5, 1, 1, 0.5, 0.5)$	CR	0.049	20.8(15.8)	10.3(7.90)	10.5(7.97)	10.6(8.08)	10.5(8.06)
$(0.5, 0.5, 1, 1, 0.4, 0.6)$	SPB	0.053	1.31(1.28)	0.664(0.621)	0.668(0.626)	0.668(0.617)	0.658(0.626)
$(0.5, 0.5, 1, 1, 0.4, 0.6)$	CR	0.048	20.9(16.1)	10.3(7.84)	8.35(6.46)	12.4(9.40)	10.2(7.85)
$(2, 2, 1, 1, 0.5, 0.5)$	SPB	0.051	1.32(1.28)	0.673(0.628)	0.659(0.621)	0.667(0.623)	0.663(0.618)
$(2, 2, 1, 1, 0.5, 0.5)$	CR	0.050	20.9(16.0)	10.4(7.94)	10.6(8.08)	10.4(7.89)	10.4(8.00)
$(2, 2, 1, 1, 0.4, 0.6)$	SPB	0.049	1.33(1.27)	0.658(0.620)	0.680(0.626)	0.673(0.625)	0.664(0.623)
$(2, 2, 1, 1, 0.4, 0.6)$	CR	0.051	20.9(16.1)	10.3(7.83)	8.42(6.39)	12.5(9.53)	10.3(7.81)

Table 4.18: Performance for SPB and complete randomization under H_0 when both covariates are continuous and t-test are used. Adjusted statistics for SPB and unadjusted statistics for complete randomization are used for data analysis.

$(\mu_1, \mu_2, \beta_1, \beta_2, p_1, p_2)$	Design	α	DIF_n	DIF_{11}	DIF_{10}	DIF_{01}	DIF_{00}
$(0.5, 0.5, 1, 1, 0.5, 0.5)$	SPB	0.050	1.33(1.27)	0.675(0.624)	0.669(0.618)	0.662(0.618)	0.670(0.629)
$(0.5, 0.5, 1, 1, 0.5, 0.5)$	CR	0.051	20.8(16.0)	10.5(8.01)	10.5(8.02)	10.4(7.98)	10.4(8.02)
$(0.5, 0.5, 1, 1, 0.4, 0.6)$	SPB	0.051	1.32(1.27)	0.678(0.631)	0.669(0.626)	0.669(0.621)	0.666(0.630)
$(0.5, 0.5, 1, 1, 0.4, 0.6)$	CR	0.047	20.9(16.2)	10.3(7.87)	8.40(6.39)	12.5(9.67)	10.3(7.77)
$(2, 2, 1, 1, 0.5, 0.5)$	SPB	0.050	1.31(1.26)	0.669(0.626)	0.657(0.622)	0.658(0.619)	0.672(0.626)
$(2, 2, 1, 1, 0.5, 0.5)$	CR	0.052	20.9(16.0)	10.3(7.91)	10.4(7.99)	10.5(8.04)	10.5(7.90)
$(2, 2, 1, 1, 0.4, 0.6)$	SPB	0.054	1.32(1.28)	0.660(0.627)	0.674(0.626)	0.670(0.623)	0.667(0.620)
$(2, 2, 1, 1, 0.4, 0.6)$	CR	0.048	21.2(16.1)	10.0(7.77)	8.41(6.36)	12.6(9.55)	10.3(7.88)

Table 4.19: Performance for Pocock and Simon's design and complete randomization under H_0 when both covariates are discrete and t-test are used. Adjusted statistics for Pocock and Simon's design and unadjusted statistics for complete randomization are used for data analysis.

$(\mu_1, \mu_2, \beta_1, \beta_2, p_1, p_2)$	Design	α	DIF_n	DIF_{11}	DIF_{10}	DIF_{01}	DIF_{00}	$DIF_{1\cdot}$	$DIF_{0\cdot}$	$DIF_{\cdot 1}$	$DIF_{\cdot 0}$
$(0.5, 0.5, 1, 1, 0.5, 0.5)$	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)
$(0.5, 0.5, 1, 1, 0.5, 0.5)$	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)
$(0.5, 0.5, 1, 1, 0.4, 0.6)$	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)
$(0.5, 0.5, 1, 1, 0.4, 0.6)$	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)
$(2, 2, 1, 1, 0.5, 0.5)$	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)
$(2, 2, 1, 1, 0.5, 0.5)$	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)
$(2, 2, 1, 1, 0.4, 0.6)$	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)
$(2, 2, 1, 1, 0.4, 0.6)$	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)

Table 4.20: Performance for Pocock and Simon's design and complete randomization under H_0 when both covariates are continuous and t-test are used. Adjusted statistics for Pocock and Simon's design and unadjusted statistics for complete randomization are used for data analysis.

$(\mu_1, \mu_2, \beta_1, \beta_2, p_1, p_2)$	Design	α	DIF_n	DIF_{11}	DIF_{10}	DIF_{01}	DIF_{00}	$DIF_{1\cdot}$	$DIF_{0\cdot}$	$DIF_{\cdot 1}$	$DIF_{\cdot 0}$
$(0.5, 0.5, 1, 1, 0.5, 0.5)$	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)
$(0.5, 0.5, 1, 1, 0.5, 0.5)$	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)
$(0.5, 0.5, 1, 1, 0.4, 0.6)$	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)
$(0.5, 0.5, 1, 1, 0.4, 0.6)$	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)
$(2, 2, 1, 1, 0.5, 0.5)$	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)
$(2, 2, 1, 1, 0.5, 0.5)$	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)
$(2, 2, 1, 1, 0.4, 0.6)$	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)
$(2, 2, 1, 1, 0.4, 0.6)$	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)

Chapter 5

Conclusions

In this dissertation, I investigated sequential monitoring of clinical trials with sample size re-estimation (SSR) under two different adaptive randomization designs, i.e., response-adaptive randomization (RAR) and covariate-adaptive randomization (CAR).

Response-adaptive randomization has been shown to have ethical and efficient advantages such as assigning more patients to the better treatment and maximizing the power of detecting the treatment differences. Its theoretical and numerical properties have been well studied. However, in order to apply RAR in real clinical trials, more research is needed. In modern clinical trials, sequential monitoring and sample size re-estimation are very popular and desirable. Clinical trial practitioners would like to combine sequential monitoring, SSR and response-adaptive randomization in one trial when considering whether to implement RAR in the trials. RAR assigns the next patient based on previous treatment assignments and responses. Therefore, the commonly used methods based on independently and identically distributed responses is not applicable any more. Moreover, sequential monitoring involve interdependent sequential test statistics. The critical step to control the type I error rate is to derive the joint distribution of the sequential test statistics. Sample size re-estimation is adaptive. Therefore, I have worked on the combination of three types of adaptive design in one trial. In this dis-

sertation, I derived the joint distribution of the proposed sequential test statistics when SSR is implemented. I also performed comprehensive numerical and simulation studies to show that my proposed method can control the type I error rate well, and enhance ethical and efficient aspects of clinical trials.

In real clinical trials, covariate-adaptive randomization designs including the stratified permuted block randomization and Pocock and Simon's design (1975) are popular randomization design in Phase III confirmatory clinical trials. As mentioned above, sequential monitoring and SSR are also very popular in practice. As a result, sequential monitoring of covariate-adaptive randomized clinical trials with SSR are very commonly used. However, theoretical investigations on this procedure is lacking in the literature. In particular, researchers realized that the type I error rate will be conservative if we do not include all the covariates used in the CAR design in the data analysis. In summary, I studied sequential monitoring of covariate-adaptive randomized clinical trials with SSR for three scenarios: 1. all the randomization covariates are used in the data analysis; 2. part of the randomization covariates are used in the data analysis; 3. none of the randomization covariates are used in the data analysis (t-test). I also theoretically showed that my method can control the type I error rate. The numerical and simulation studies supported my theoretical findings.

There are many directions for future research. I have mentioned a few in previous chapters. Here I emphasize one direction from the point of view of the adaptive randomization designs. Clinical trials often involve various covariates since the heterogeneity of patients' responses to a treatment is well-accepted as the development of Bioinformatics. At the same time, the ethical and efficient considerations are expected to be dealt with in clinical trials. Zhang et al. (2007) proposed covariate-adjusted response-adaptive randomization in order to preserve the advantages of RAR while taking into account the heterogeneity of patients' responses to a treatment. However, this design requires

quite difficult theoretical foundations. As a result, lots of fundamental properties are unclear for this design. For example, when covariates are continuous, how can we define the so-called allocation proportion. Even without SSR, how can we sequentially monitor the CARA design while controlling the type I error rate. Traditional spending function methods are based on standard Brownian motion with a fundamental assumption that the increment of the monitoring statistic is independent. Brownian motion have been used in many fields such as in dynamic systems and economics (Hu et al., 2003; Jumarie, 2006). Brownian motion provided a lot of useful theoretical results in monitoring clinical trials (Lan and Wittes, 1988; Davis and Hardy, 1990, 1994). Although these methods were derived under several assumptions, it is a common feature that the test statistic forms a Brownian motion over the information time (Lachin, 2005). However in practice, the assumptions may not be satisfied, since patients are followed for a long time period and the test statistic is formed with aggregations from a group of patients. Fractional Brownian motion is a model to deal with the long-memory stochastic processes due to aggregation. Lai (2010) studies the boundaries under factional Brownian motion for five α spending functions. A more comprehensive method to decide whether stopping or continuing the trials was provided based on the new results. It is also worth studying this scenario. I leave all these for future research.

REFERENCE

[1] Aickin M. Effect of design-adaptive allocation on inference for a regression parameter: Two-group, single-covariate and double-covariate cases. Statistics & Probability Letters. 2009 Jan 1;79(1):16-20.

[2] Andersen J, Faries D, Ramura R. A randomized play-the-winner design for multi-arm clinical trials. Communications in Statistics-Theory and Methods. 1994 Jan 1;23(2):309-23.

[3] Anderson H, Hopwood P, Stephens RJ, Thatcher N, Cottier B, Nicholson M, Milroy R, Maughan TS, Falk SJ, Bond MG, Burt PA. Gemcitabine plus best supportive care (BSC) vs BSC in inoperable non-small cell lung cancera randomized trial with quality of life as the primary outcome. British Journal of Cancer. 2000 Aug;83(4):447.

[4] Anderson KM. Optimal spending functions for asymmetric group sequential designs. Biometrical Journal: Journal of Mathematical Methods in Biosciences. 2007 Jun;49(3):337-45.

[5] Armitage P. Sequential Medical Trials Blackwell.1975.

[6] Athreya KB. On a characteristic property of P*ó*lya's urn. Studia Sci. Math. Hungar. 1969;4:31-5.

[7] Athreya KB, Karlin S. Limit theorems for the split times of branching processes. Journal of Mathematics and Mechanics. 1967 Jan 1;17(3):257-77.

[8] Athreya KB, Karlin S. Embedding of urn schemes into continuous time Markov branching processes and related limit theorems. The Annals of Mathematical Statistics. 1968 Dec 1;39(6):1801-17.

[9] Bai ZD, Hu F, Shen L. An adaptive design for multi-arm clinical trials. Journal of Multivariate Analysis. 2002 Apr 1;81(1):1-8.

[10] Bai ZD, Hu F, Rosenberger WF. Asymptotic properties of adaptive designs for clinical trials with delayed response. In Advances In Statistics 2008 (pp. 263-280).

[11] Bartlett RH, Roloff DW, Cornell RG, Andrews AF, Dillon PW, Zwischenberger JB. Extracorporeal circulation in neonatal respiratory failure: a prospective randomized study. Pediatrics. 1985 Oct 1;76(4):479-87.

[12] Bauer P, Köhne K. Evaluation of experiments with adaptive interim analyses. Biometrics. 1994 Dec 1:1029-41.

[13] Birkett NJ. Adaptive allocation in randomized controlled trials. Controlled Clinical Trials. 1985 Jun 1;6(2):146-55.

[14] Chang M. Adaptive design theory and implementation using SAS and R. Chapman and Hall/CRC; 2014 Dec 1.

[15] Chen TT, Ng TH. Optimal flexible designs in phase II clinical trials. Statistics in Medicine. 1998 Oct 30;17(20):2301-12.

[16] Chen TT. Optimal three-stage designs for phase II cancer clinical trials. Statistics in Medicine. 1997 Dec 15;16(23):2701-11.

[17] Chow SC, Chang M. Adaptive design methods in clinical trials. CRC press; 2011 Dec 1.

[18] Chow SC, Shao J, Wang H. Sample size calculations in clinical research. Chapman and Hall/CRC; 2008.

[19] Connor EM, Sperling RS, Gelber R, Kiselev P, Scott G, O'sullivan MJ, VanDyke R, Bey M, Shearer W, Jacobson RL, Jimenez E. Reduction of maternal-infant transmission of human immunodeficiency virus type 1 with zidovudine treatment. New England Journal of Medicine. 1994 Nov 3;331(18):1173-80.

[20] Cui L, Hung HM, Wang SJ. Modification of sample size in group sequential clinical trials. Biometrics. 1999 Sep 1;55(3):853-7.

[21] Davis BR, Hardy RJ. Upper bounds for type I and type II error rates in conditional power calculations. Communications in Statistics-Theory and Methods. 1990 Jan 1;19(10):3571-84.

[22] Davis BR, Hardy RJ. Data monitoring in clinical trials: the case for stochastic curtailment. Journal of Clinical Epidemiology. 1994 Sep 1;47(9):1033-42.

[23] Denne JS. Sequential procedures for sample size estimation (Doctoral dissertation, University of Bath) 1996.

[24] DeMets DL. Futility approaches to interim monitoring by data monitoring committees. Clinical Trials. 2006 Dec;3(6):522-9.

[25] Durham SD, Yu KF. Randomized play-the-leader rules for sequential sampling from two populations. Probability in the Engineering and Informational Sciences. 1990 Jul;4(3):355-67.

[26] Durham SD, Flournoy N, Li W. A sequential design for maximizing the probability of a favourable response. Canadian Journal of Statistics. 1998 Sep;26(3):479-95.

[27] Efron B. Forcing a sequential experiment to be balanced. Biometrika. 1971 Dec 1;58(3):403-17.

[28] Eggenberger F, Pólya G. Über die statistik verketteter vorgänge. ZAMM-Journal of Applied Mathematics and Mechanics/Zeitschrift für Angewandte Mathematik und Mechanik. 1923;3(4):279-89.

[29] Eisele JR. The doubly adaptive biased coin design for sequential clinical trials. Journal of Statistical Planning and Inference. 1994 Feb 1;38(2):249-61.

[30] Eisele JR, Woodroofe MB. Central limit theorems for doubly adaptive biased coin designs. The Annals of Statistics. 1995 Feb 1:234-54.

[31] Ensign LG, Gehan EA, Kamen DS, Thall PF. An optimal three-stage design for phase II clinical trials. Statistics in Medicine. 1994 Sep 15;13(17):1727-36.

[32] FDA. Draft Guidance for Industry: Adaptive Design Clinical Trials for Drugs and Biologics. 2010.

[33] Feinstein AR, Landis JR. The role of prognostic stratification in preventing the bias permitted by random allocation of treatment. Journal of Chronic Diseases. 1976 Apr 1;29(4):277-84.

[34] Forsythe AB. Validity and power of tests when groups have been balanced for prognostic factors. Computational Statistics & Data Analysis. 1987 Jan 1;5(3):193-200.

[35] Friedman B. A simple urn model. Communications on Pure and Applied Mathematics. 1949 Mar;2(1):59-70.

[36] Friedman LM, Furberg C, DeMets DL, Reboussin D, Granger CB. Fundamentals of Clinical Trials. Springer-Verlag; 2015 Aug 27.

[37] Gordon Lan KK, DeMets DL. Discrete sequential boundaries for clinical trials. Biometrika. 1983 Dec 1;70(3):659-63.

[38] Lawrence Gould A, Shih WJ. Sample size re-estimation without unblinding for normally distributed outcomes with unknown variance. Communications in Statistics-Theory and Methods. 1992 Jan 1;21(10):2833-53.

[39] Gould AL. Interim analyses for monitoring clinical trials that do not materially affect the type I error rate. Statistics in medicine. 1992;11(1):55-66.

[40] Gould AL, Shih WJ. Modifying the design of ongoing trials without unblinding. Statistics in Medicine. 1998 Jan 15;17(1):89-100.

[41] Green SB, Byar DP. The effect of stratified randomization on size and power of statistical tests in clinical trials. Journal of Clinical Epidemiology. 1978 Jan 1;31(6):445-54.

[42] Gridelli C, Perrone F, Gallo C, Cigolari S, Rossi A, Piantedosi F, Barbera S, Ferra F, Piazza E, Rosetti F, Clerici M. Chemotherapy for elderly patients with advanced non-small-cell lung cancer: the Multicenter Italian Lung Cancer in the Elderly Study (MILES) phase III randomized trial. Journal of the National Cancer Institute. 2003 Mar 5;95(5):362-72.

[43] Herson J, Wittes J. The use of interim analysis for sample size adjustment. Drug Information Journal. 1993 Jul;27(3):753-60.

[44] Gwise T, Hu F, Hu J. Optimal biased coins for two-arm clinical trials. Statistics and its Interface. 2008;1(1):125-35.

[45] Haybittle JL. Repeated assessment of results in clinical trials of cancer treatment. The British journal of radiology. 1971 Oct;44(526):793-7.

[46] Heritier S, Gebski V, Pillai A. Dynamic balancing randomization in controlled clinical trials. Statistics in medicine. 2005 Dec 30;24(24):3729-41.

[47] Herson J, Wittes J. The use of interim analysis for sample size adjustment. Drug Information Journal. 1993 Jul;27(3):753-60.

[48] Hu Y, Hu F. Asymptotic properties of covariate-adaptive randomization. The Annals of Statistics. 2012;40(3):1794-815.

[49] Hu Y, Øksendal B, Sulem A. Optimal consumption and portfolio in a Black-Scholes market driven by fractional Brownian motion. Infinite Dimensional Analysis, Quantum Probability and Related Topics. 2003 Dec;6(04):519-36.

[50] Hu F, Rosenberger WF. Optimality, variability, power: evaluating response-adaptive randomization procedures for treatment comparisons. Journal of the American Statistical Association. 2003 Sep 1;98(463):671-8.

[51] Hu F, Rosenberger WF. The theory of response-adaptive randomization in clinical trials. John Wiley & Sons; 2006 Sep 29.

[52] Hu F, Rosenberger WF, Zhang LX. Asymptotically best response-adaptive randomization procedures. Journal of Statistical Planning and Inference. 2006 Jun 1;136(6):1911-22.

[53] Hu F, Zhang LX. Asymptotic properties of doubly adaptive biased coin designs for multitreatment clinical trials. The Annals of Statistics. 2004;32(1):268-301.

[54] Hu F, Zhang LX. Asymptotic normality of urn models for clinical trials with delayed response. Bernoulli. 2004;10(3):447-63.

[55] Hu F, Zhang LX, Cheung SH, Chan WS. Doubly adaptive biased coin designs with delayed responses. Canadian Journal of Statistics. 2008 Dec;36(4):541-59.

[56] Hu F, Zhang LX, He X. Efficient randomized-adaptive designs. The Annals of Statistics. 2009 Oct 1:2543-60.

[57] Ivanova A, Rosenberger WF. A comparison of urn designs for randomized clinical trials of $K > 2$ treatments. Journal of biopharmaceutical statistics. 2000 Feb 15;10(1):93-107.

[58] Ivanova A, Flournoy N. A birth and death urn for ternary outcomes: stochastic processes applied to urn models. Probability and Statistical Models with Applications. 2001:583-600.

[59] Ivanova A. A play-the-winner-type urn design with reduced variability. Metrika. 2003 Aug 1;58(1):1-3.

[60] Iacono AT, Johnson BA, Grgurich WF, Youssef JG, Corcoran TE, Seiler DA, Dauber JH, Smaldone GC, Zeevi A, Yousem SA, Fung JJ. A randomized trial of inhaled cyclosporine in lung-transplant recipients. New England Journal of Medicine. 2006 Jan 12;354(2):141-50.

[61] Jacob G, Comparetti PM, Neugschwandtner M, Kruegel C, Vigna G. A static, packer-agnostic filter to detect similar malware samples. In International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment 2012 Jul 26 (pp. 102-122). Springer, Berlin, Heidelberg.

[62] Jennison C, Turnbull BW. Group-sequential analysis incorporating covariate information. Journal of the American Statistical Association. 1997 Dec 1;92(440):1330-41.

[63] Jennison C, Turnbull BW. Group Sequential Methods With Applications to Clinical Trials. 2000 Boca Raton. FL Chapman & Hall/CRC.

[64] Jumarie G. New stochastic fractional models for Malthusian growth, the Poissonian birth process and optimal management of populations. Mathematical and Computer Modelling. 2006 Aug 1;44(3-4):231-54.

[65] Krueger GG, Langley RG, Leonardi C, Yeilding N, Guzzo C, Wang Y, Dooley LT, Lebwohl M. A human interleukin-12/23 monoclonal antibody for the treatment of psoriasis. New England Journal of Medicine. 2007 Feb 8;356(6):580-92.

[66] Lachin JM. A review of methods for futility stopping based on conditional power. Statistics in Medicine. 2005 Sep 30;24(18):2747-64.

[67] Lai HL, Chen CJ, Peng TC, Chang FM, Hsieh ML, Huang HY, Chang SC. Randomized controlled trial of music during kangaroo care on maternal state anxiety and preterm infants' responses. International journal of nursing studies. 2006 Feb 1;43(2):139-46.

[68] Lai D. Group sequential tests under fractional Brownian motion in monitoring clinical trials. Statistical Methods and Applications. 2010 Jun 1;19(2):277-86.

[69] Lai D. Sample size determination for group sequential test under fractional Brownian motion. Electronic Journal of Statistics. 2013;7:1957-67.

[70] Gordon Lan KK, DeMets DL. Discrete sequential boundaries for clinical trials. Biometrika. 1983 Dec 1;70(3):659-63.

[71] Lan KG, Wittes J. The B-value: a tool for monitoring data. Biometrics. 1988 Jun 1:579-85.

[72] Lehmacher W, Wassmer G. Adaptive sample size calculations in group sequential trials. Biometrics. 1999 Dec 1;55(4):1286-90.

[73] Lehmann EL. Elements of large-sample theory. Springer Science & Business Media; 2004 Aug 27.

[74] Ma W, Hu F, Zhang L. Testing hypotheses of covariate-adaptive randomized clinical trials. Journal of the American Statistical Association. 2015 Apr 3;110(510):669-80.

[75] McEntegart DJ. The pursuit of balance using stratified and dynamic randomization techniques: an overview. Drug Information Journal. 2003 Jul;37(3):293-308.

[76] Mehta CR, Pocock SJ. Adaptive increase in sample size when interim results are promising: a practical guide with examples. Statistics in Medicine. 2011 Dec 10;30(28):3267-84.

[77] Molander A, Warfvinge J, Reit C, Kvist T. Clinical and radiographic evaluation of one-and two-visit endodontic treatment of asymptomatic necrotic teeth with apical periodontitis: a randomized clinical trial. Journal of endodontics. 2007 Oct 1;33(10):1145-8.

[78] Neyman J. On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection. Journal of the Royal Statistical Society. 1934 Jan 1;97(4):558-625.

[79] Nordle O, Brantmark BO. A self-adjusting randomization plan for allocation of patients into two treatment groups. Clinical Pharmacology & Therapeutics. 1977 Dec;22(6):825-30.

[80] O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. Biometrics. 1979 Sep 1:549-56.

[81] Ohtori S, Miyagi M, Eguchi Y, Inoue G, Orita S, Ochiai N, Kishida S, Kuniyoshi K, Nakamura J, Aoki Y, Ishikawa T. Epidural administration of spinal nerves with the tumor necrosis factor-alpha inhibitor, etanercept, compared with dexamethasone for treatment of sciatica in patients with lumbar spinal stenosis: a prospective randomized study. Spine. 2012 Mar 15;37(6):439-44.

[82] Peto R, Pike M, Armitage P, Breslow NE, Cox DR, Howard SV, Mantel N, McPherson K, Peto J, Smith PG. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. I. Introduction and design. British journal of cancer. 1976 Dec;34(6):585.

[83] Pocock SJ, Simon R. Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. Biometrics. 1975 Mar 1:103-15.

[84] Pocock SJ. Group sequential methods in the design and analysis of clinical trials. Biometrika. 1977 Aug 1;64(2):191-9.

[85] Proschan MA, Hunsberger SA. Designed extension of studies based on conditional power. Biometrics. 1995 Dec 1:1315-24.

[86] Proschan MA. Two-stage sample size re-estimation based on a nuisance parameter: a review. Journal of biopharmaceutical statistics. 2005 Jul 1;15(4):559-74.

[87] Proschan MA, Lan KG, Wittes JT. Statistical monitoring of clinical trials: a unified approach. Springer Science & Business Media; 2006 Dec 31.

[88] Robbins H. Some aspects of the sequential design of experiments. Bull. Amer. Math. Soc. 1952 58 527?535.

[89] Rosenberger WF, Lachin JM. Randomization in clinical trials: theory and practice. John Wiley & Sons; 2015 Nov 23.

[90] Rosenberger WF, Stallard N, Ivanova A, Harper CN, Ricks ML. Optimal adaptive designs for binary response trials. Biometrics. 2001 Sep;57(3):909-13.

[91] Rosenberger WF, Sverdlov O. Handling covariates in the design of clinical trials. Statistical Science. 2008 Aug 1:404-19.

[92] Rout CC, Rocke DA, Levin J, Gouws E, Reddy D. A reevaluation of the role of crystalloid preload in the prevention of hypotension associated with spinal anesthesia for elective cesarean section. Anesthesiology. 1993 Aug;79(2):262-9.

[93] Rothwell PM. External validity of randomised controlled trials:"to whom do the results of this trial apply?". The Lancet. 2005 Jan 1;365(9453):82-93.

[94] Sargent DJ, Goldberg RM. A flexible design for multiple armed screening trials. Statistics in Medicine. 2001 Apr 15;20(7):1051-60.

[95] Sertkaya, A., Wong, H. H., Jessup, A., & Beleche, T. (2016). Key cost drivers of pharmaceutical clinical trials in the United States. Clinical Trials, 13(2), 117-126.

[96] Shao J, Yu X, Zhong B. A theory for testing hypotheses under covariate-adaptive randomization. Biometrika. 2010 Apr 30;97(2):347-60.

[97] Shao J, Yu X. Validity of Tests under Covariate-Adaptive Biased Coin Randomization and Generalized Linear Models. Biometrics. 2013 Dec;69(4):960-9.

[98] Shih WJ. Sample size re-estimation in clinical trials. In biopharmaceutical sequential statistical applications (Ed., K.E. Peace), New York: Marcel Dekker, 1992 285-301.

[99] Shih WJ, Aisner J. Statistical design and analysis of clinical trials: principles and methods. Chapman and Hall/CRC; 2015 Jul 28.

[100] Signorini DF, Leung O, Simes RJ, Beller E, Gebski VJ, Callaghan T. Dynamic balanced randomization for clinical trials. Statistics in medicine. 1993 Dec 30;12(24):2343-50.

[101] Simon R. Optimal two-stage designs for phase II clinical trials. Contemporary Clinical Trials. 1989 Mar 1;10(1):1-0.

[102] Smythe RT. Central limit theorems for urn models. Stochastic Processes and their Applications. 1996 Dec 13;65(1):115-37.

[103] Tamura RN, Faries DE, Andersen JS, Heiligenstein JH. A case study of an adaptive clinical trial in the treatment of out-patients with depressive disorder. Journal of the American Statistical Association. 1994 Sep 1;89(427):768-76.

[104] Taves DR. The use of minimization in clinical trials. Contemporary Clinical Trials. 2010 Mar 1;31(2):180-4.

[105] Taves DR. Minimization: a new method of assigning patients to treatment and control groups. Clinical Pharmacology & Therapeutics. 1974 May 1;15(5):443-53.

[106] Thompson WR. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. Biometrika. 1933 Dec 1;25(3/4):285-94.

[107] Tu D, Shalay K, Pater J. Adjustment of treatment effect for covariates in clinical trials: statistical and regulatory issues. Drug Information Journal. 2000 Apr;34(2):511-23.

[108] Tymofyeyev Y, Rosenberger WF, Hu F. Implementing optimal allocation in sequential binary response experiments. Journal of the American Statistical Association. 2007 Mar 1;102(477):224-34.

[109] Wald A. Sequential analysis. 1947 John Wiley and Sons Inc., New York.

[110] Wang SK, Tsiatis AA. Approximately optimal one-parameter boundaries for group sequential trials. Biometrics. 1987 Mar 1:193-9.

[111] Wei LJ. An application of an urn model to the design of sequential controlled clinical trials. Journal of the American Statistical Association. 1978 Sep 1;73(363):559-63.

[112] Wei LJ, Durham S. The randomized play-the-winner rule in medical trials. Journal of the American Statistical Association. 1978 Dec 1;73(364):840-3.

[113] Wei LJ. The generalized Pólya's urn design for sequential medical trials. The Annals of Statistics. 1979 Mar 1:291-6.

[114] Whitehead J. The design and analysis of sequential clinical trials. John Wiley & Sons; 1997 Aug 4.

[115] Wittes J, Brittain E. The role of internal pilot studies in increasing the efficiency of clinical trials. Statistics in Medicine. 1990 Jan;9(1-2):65-72.

[116] Wu MC, Lan KG. Sequential monitoring for comparison of changes in a response variable in clinical studies. Biometrics. 1992 Sep 1:765-79.

[117] Zhang LX, Hu F, Cheung SH. Asymptotic theorems of sequential estimation-adjusted urn models. The Annals of Applied Probability. 2006;16(1):340-69.

[118] Zhang LX, Hu F, Cheung SH, Chan WS. Asymptotic properties of covariate-adjusted response-adaptive designs. The Annals of Statistics. 2007;35(3):1166-82.

[119] Zelen M. Play the winner rule and the controlled clinical trial. Journal of the American Statistical Association. 1969 Mar 1;64(325):131-46.

[120] Zhu H, Hu F. Interim analysis of clinical trials based on urn models. Canadian Journal of Statistics. 2012 Sep;40(3):550-68.

[121] Zhu H, Hu F. Sequential monitoring of covariate-adaptive randomized clinical trials. Statistica Sinica. 2018 in press.

APPENDIX

Proof of Theorem 2.1, 3.1, 4.1 and 4.2: Note that before involving sample size re-estimation, the sequential statistics Z_t defined in different chapters in this dissertation are the same as those in Zhu and Hu (2012) (urn models) and Zhu and Hu (2018) (CAR). Under the regularity conditions of the corresponding chapters, Theorem 2 of Zhu and Hu (2012) proved that under H_0, $B_t = \sqrt{t}Z_t$ in Chapter 2 of this dissertation converges to a standard Brownian motion in distribution. That is, $\{Z_{t_1}, ..., Z_{t_K}\}$ is multivariate normal; $EZ_{t_i} = 0$; and $Cov(Z_{t_i}, Z_{t_j}) = \sqrt{[nt_i]/[nt_j]}$, $0 \leq t_i \leq t_j \leq 1$. Theorem 2.1 of Zhu and Hu (2018) proved that under H_0, $B_t = \sqrt{t}Z_t$ in Chapter 3, $B_t = \sqrt{t}Z_t^{adj}$ and $B_t = \sqrt{t}Z_t^{adj'}$ in Chapter 4 of this dissertation converge to a standard Brownian motion in distribution. That is, $\{Z_{t_1}^{adj}, ..., Z_{t_K}^{adj}\}$ is multivariate normal; $EZ_{t_i}^{adj} = 0$; and $Cov(Z_{t_i}^{adj}, Z_{t_j}^{adj}) = \sqrt{t_i/t_j}$, $0 \leq t_i \leq t_j \leq 1$. To save space, I use the notation Z_t to represent Z_t, Z_t^{adj} and $Z_t^{adj'}$ hereafter to offer a unified proof for Theorem 2.1, 3.1, 4.1 and 4.2.

Note that in this dissertation, my test statistics U_t are defined as follows,

$$U_t = Z_t, \text{ if } t \leq t_L;$$

and

$$U_t = w_t^{1/2} Z_{t_L} + (1 - w_t)^{1/2} \frac{B_{b(t-t_L)+t_L} - B_{t_L}}{\{b(t - t_L)\}^{1/2}}, \text{ if } t_L \leq t \leq 1,$$

where $w_t = t_L/t$ and $b = (w - t_L)/(1 - t_L)$. Therefore, based on the conclusion of Zhu and Hu (2012) and Zhu and Hu (2018), we only need to prove that the joint distribution of $(U_{t_1}, \ldots, U_{t_K})$ is the same as that of $(Z_{t_1}, \ldots, Z_{t_K})$ under H_0. Firstly, U_t is the linear combination of Z_t, so $(U_{t_1}, \ldots, U_{t_K})$ also follows multivariate normal distribution. Next, I will prove that $(U_{t_1}, \ldots, U_{t_K})$ have the same mean, variance and covariance as $(Z_{t_1}, \ldots, Z_{t_K})$.

It is clear, for $t \leq t_L$, we have $U_t = Z_t$ by definition.

For $t_L \leq t \leq 1$, we have

$$\begin{aligned}
E(U_t|Z_{t_L}) &= E\left(w_t^{1/2} Z_{t_L} + (1-w_t)^{1/2} \frac{B_{b(t-t_L)+t_L} - B_{t_L}}{\{b(t-t_L)\}^{1/2}} \Big| Z_{t_L}\right) \\
&= E\left(w_t^{1/2} B_{t_L}/\sqrt{t_L} + (1-w_t)^{1/2} \frac{B_{b(t-t_L)+t_L} - B_{t_L}}{\{b(t-t_L)\}^{1/2}} \Big| B_{t_L}\right) \\
&= E\left(w_t^{1/2} B_{t_L}/\sqrt{t_L} \Big| B_{t_L}\right) \\
&= w_t^{1/2} B_{t_L}/\sqrt{t_L} \\
&= w_t^{1/2} Z_{t_L}.
\end{aligned}$$

$$E(U_t) = E(E(U_t|Z_{t_L})) = E(w_t^{1/2} Z_{t_L}) = 0.$$

$$\begin{aligned}
Var(U_t|Z_{t_L}) &= Var\left(w_t^{1/2} B_{t_L}/\sqrt{t_L} + (1-w_t)^{1/2} \frac{B_{b(t-t_L)+t_L} - B_{t_L}}{\{b(t-t_L)\}^{1/2}} \Big| B_{t_L}\right) \\
&= Var\left((1-w_t)^{1/2} \frac{B_{b(t-t_L)+t_L} - B_{t_L}}{\{b(t-t_L)\}^{1/2}} \Big| B_{t_L}\right) \\
&= \frac{1-w_t}{b(t-t_L)} (b(t-t_L) + t_L - t_L) \\
&= 1 - w_t.
\end{aligned}$$

$$\begin{aligned}
Var(U_t) &= E(Var(U_t|Z_{t_L})) + Var(E(U_t|Z_{t_L})) \\
&= E(1-w_t) + Var(w_t^{1/2} Z_{t_L}) \\
&= 1 - w_t + w_t \\
&= 1.
\end{aligned}$$

When $t_L < t_1 < t_2 < 1$,

$$
\begin{aligned}
Cov(U_{t_1}, U_{t_2}|Z_{t_L}) &= Cov\left((1-w_{t_1})^{1/2}\frac{B_{b(t_1-t_L)+t_L} - B_{t_L}}{\{b(t_1-t_L)\}^{1/2}}, (1-w_{t_2})^{1/2}\frac{B_{b(t_2-t_L)+t_L} - B_{t_L}}{\{b(t_2-t_L)\}^{1/2}}\Big|B_{t_L}\right) \\
&= \frac{(1-w_{t_1})^{1/2}(1-w_{t_2})^{1/2}}{\{b(t_1-t_L)\}^{1/2}\{b(t_2-t_L)\}^{1/2}} Cov\left(B_{b(t_1-t_L)+t_L} - B_{t_L}, B_{b(t_2-t_L)+t_L} - B_{t_L}\Big|B_{t_L}\right) \\
&= \frac{1}{b\sqrt{t_1 t_2}} Cov\big(B_{b(t_1-t_L)+t_L} - B_{t_L}, \\
&[B_{b(t_1-t_L)+t_L} - B_{t_L}] + [B_{b(t_2-t_L)+t_L} - B_{b(t_1-t_L)+t_L}]\big|B_{t_L}\big) \\
&= \frac{1}{b\sqrt{t_1 t_2}} Var\big(B_{b(t_1-t_L)+t_L} - B_{t_L}\big|B_{t_L}\big) + \\
&\frac{1}{b\sqrt{t_1 t_2}} Cov\big(B_{b(t_1-t_L)+t_L} - B_{t_L}, [B_{b(t_2-t_L)+t_L} - B_{b(t_1-t_L)+t_L}]\big|B_{t_L}\big) \\
&= \frac{1}{b\sqrt{t_1 t_2}} Var\big(B_{b(t_1-t_L)+t_L} - B_{t_L}\big|B_{t_L}\big) \\
&= \frac{1}{b\sqrt{t_1 t_2}} (b(t_1-t_L) + t_L - t_L) \\
&= \frac{t_1 - t_L}{\sqrt{t_1 t_2}}.
\end{aligned}
$$

$$
\begin{aligned}
Cov(U_{t_1}, U_{t_2}) &= E(Cov(U_{t_1}, U_{t_2}|Z_{t_L}) + Cov(E(U_{t_1}|Z_{t_L}), E(U_{t_2}|Z_{t_L})) \\
&= E(\frac{t_1 - t_L}{\sqrt{t_1 t_2}}) + Cov(w_{t_1}^{1/2} Z_{t_L}, w_{t_2}^{1/2} Z_{t_L}) \\
&= \frac{t_1 - t_L}{\sqrt{t_1 t_2}} + w_{t_1}^{1/2} w_{t_2}^{1/2} Var(Z_{t_L}) \\
&= \frac{t_1 - t_L}{\sqrt{t_1 t_2}} + (t_L/t_1)^{1/2}(t_L/t_2)^{1/2} \\
&= (t_1/t_2)^{1/2}.
\end{aligned}
$$

Therefore, we have the joint distribution of $(U_{t_1}, \ldots, U_{t_K})$ is the same as that of $(Z_{t_1}, \ldots, Z_{t_K})$ under H_0.

R code for Example 1 of Chapter 2

```
simulation=function(pa,pb,n1,n2,n3_0,m,c1,c2,c3,ssrfl)
{
#pa pb are success rates for treatment A and B, respectively
#n1,n2,n3_0 are the originally planned sample size for the three stages
#m is the number of replications
#c1,c2,c3 are the critical values
#ssrfl is the indicator for whether to implement SSR.
#total sample size originally planned
ntotal=n1+n2+n3_0
#max number for 3rd stage after SSR
nmax = 500
#desired conditional power value
pcut=0.9
#t1 and t2 are the information times for the first two looks
t1=0.2
t2=0.5
#total number of failures
failure=NULL
#failure rates
failureratio=NULL
#number of rejection of H_0 out of m replications
number=0
#number of cases when SSR was implemented
numofssr=0
#rho1 and rho2 are actual allocation proportions for the two treatments, respectively
rho1<-NULL
rho2<-NULL
#urn1 and urn2 are the urn compositions (number of balls of type 1 and 2)
urn1<-NULL
urn2<-NULL
#number of rejection at first/second/third look
reject1=0
reject2=0
reject3=0
#final sample size
SS=NULL
#if SSR will be done, the increase of sample size
SSplus=NULL
for  ( i in 1:m ){
        #number of type 1 ball in the urn, initial numbers are 5 for both treatments
        ball1=5
        ball2=5
        #responses of patients in treatment 1/2
        xx1<-NULL
        xx2<-NULL
        #number of patients in treatment 1/2
        N1=0
        N2=0

        for (j in 1:n1){
                Rho1=ball1/(ball1+ball2)
```

```
        #Rho1 is the probability that drawing type 1
    x<-runif(1,0,1)
        if (x>=0 & x<Rho1) {
                N1<-N1+1
    new=rbinom(1,1,pa)
                xx1<-c(xx1,new)
        }
        if (x>=Rho1 & x<=1) {
                N2<-N2+1
    new=rbinom(1,1,pb)
                xx2<-c(xx2,new)
        }
    p1hat=(sum(xx1)+1)/(N1+1)
    p2hat=(sum(xx2)+1)/(N2+1)
    ball1=ball1+sqrt(p1hat)
    ball2=ball2+sqrt(p2hat)
        }
p1hat=(sum(xx1)+1)/(N1+1)
p2hat=(sum(xx2)+1)/(N2+1)

stat=abs((p1hat-p2hat)/sqrt(p1hat*(1-p1hat)/N1+p2hat*(1-p2hat)/N2)) #statistic
if (stat>c1) {
number=number+1
reject1=reject1+1
} else {
for (j in 1:n2){
        Rho1=ball1/(ball1+ball2)
        #probability that drawing type 1
    x<-runif(1,0,1)
        if (x>=0 & x<Rho1) {
                N1<-N1+1
    new=rbinom(1,1,pa)
                xx1<-c(xx1,new)
        }
        if (x>=Rho1 & x<=1) {
                N2<-N2+1
    new=rbinom(1,1,pb)
                xx2<-c(xx2,new)
        }
    p1hat=(sum(xx1)+1)/(N1+1)
    p2hat=(sum(xx2)+1)/(N2+1)
    ball1=ball1+sqrt(p1hat)
    ball2=ball2+sqrt(p2hat)
        }
p1hat=(sum(xx1)+1)/(N1+1)
p2hat=(sum(xx2)+1)/(N2+1)

stat2=abs((p1hat-p2hat)/sqrt(p1hat*(1-p1hat)/N1+p2hat*(1-p2hat)/N2))
if (stat2>c2) {
number=number+1
reject2=reject2+1
} else {
 #indicator of whether SSR will be implemented
```

```
    cpfl = FALSE
    #treatment effect Delta in this dissertation
    mu_1 = (mean(xx2)-mean(xx1))/sqrt(mean(xx1)*(1-mean(xx1))/(N1/(N1+N2))
           +mean(xx2)*(1-mean(xx2))/(N2/(N1+N2)))
  #conditional power
    cp_1 =1-pnorm((c3 - stat2*sqrt(t2)- mu_1*sqrt(ntotal)*(1-t2))/sqrt(1-t2) )
    if(0.01<cp_1 & cp_1<pcut) cpfl = TRUE

    fx = function(ntotal, pcut0 = pcut ){
         1-pnorm( (c3 - stat2*sqrt(t2)- mu_1*sqrt(ntotal)*(1-t2))/sqrt(1-t2) )- pcut0
  }
    if(ssrfl & cpfl){
      #sample size needed for the 3rd stage
      ncp = floor(uniroot(fx,c(n3_0,1000000))$root) - n1 - n2
      #final sample size after SSR following the rule of the dissertation
      n3 = min(nmax, max(n3_0, ncp))
      SSplus=c(SSplus,n3-n3_0)
    } else {
      n3=n3_0
    }
  for (j in 1:n3){
#probability that drawing type 1
         Rho1=ball1/(ball1+ball2)
      x<-runif(1,0,1)
         if (x>=0 & x<Rho1) {
                N1<-N1+1
      new=rbinom(1,1,pa)
                xx1<-c(xx1,new)
         }
         if (x>=Rho1 & x<=1) {
                N2<-N2+1
      new=rbinom(1,1,pb)
                xx2<-c(xx2,new)
         }
      p1hat=(sum(xx1)+1)/(N1+1)
      p2hat=(sum(xx2)+1)/(N2+1)
      ball1=ball1+sqrt(p1hat)
      ball2=ball2+sqrt(p2hat)
         }
  b = n3/n3_0
  stat30=(mean(xx2)-mean(xx1))/sqrt(mean(xx1)*(1-mean(xx1))/N1+mean(xx2)*(1-mean(xx2))/N2)
  #the third test statistic U_t
  stat3 = stat2*sqrt((n1+n2)/(n1+n2+n3_0))+
          sqrt((n1+n2+n3)/(n1+n2+n3_0))*
          (mean(xx2)-mean(xx1))/sqrt(mean(xx1)*(1-mean(xx1))/N1+mean(xx2)*(1-mean(xx2))/N2)
          *sqrt(1-(n1+n2)/(n1+n2+n3_0))
          /sqrt(b*(n3_0/(n1+n2+n3_0)))
          - stat2*sqrt((n1+n2)/(n1+n2+n3_0))
          *sqrt(1-(n1+n2)/(n1+n2+n3_0))
          /sqrt(b*(n3_0/(n1+n2+n3_0)))
  if(ssrfl & cpfl){
    stat = stat3
  }else{
```

```
            stat = stat30
          }
          if (stat>c3) {
          number=number+1
          reject3=reject3+1
          }
        }
        }
        rho1=c(rho1,N1/(N1+N2))
        rho2=c(rho2,N2/(N1+N2))
        urn1=c(urn1,ball1/(ball1+ball2))
        urn2=c(urn2,ball2/(ball1+ball2))
        SS=c(SS,N1+N2)
        failure=c(failure,length(xx1)+length(xx2)-sum(xx1)-sum(xx2))
        failureratio=c(failureratio,(length(xx1)+length(xx2)-sum(xx1)-sum(xx2))/(length(xx1)+length(xx2)))
}
result=c(number/m,mean(rho1),sd(rho1),mean(urn1),sd(urn1),
mean(SS),sd(SS),mean(failure),sd(failure),mean(failureratio),sd(failureratio))

rm(list = ls(all = TRUE))

}
```