2016

Response Adaptive Randomization using Surrogate and Primary Endpoints

Hui Wang

Response Adaptive Randomization using Surrogate and Primary Endpoints

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University.

by

Hui Wang

Director: Nitai Mukhopadhyay, Ph.D., Associate Professor, Department of Biostatistics

Department of Biostatistics

Virginia Commonwealth University

Richmond, Virginia

August, 2016

Acknowledgements

First of all, I would like to express my deepest appreciation to my thesis advisor, Dr. Nitai D. Mukhopadhyayi for his encouragement and persistent help throughout the course of my thesis. I thank him for the systematic guidance and great effort he put into training me in the scientific field.

I would also like to thank my committee members, Dr. Donna McClish, Dr. Roy Sabo, Dr. Amy Pakyz, and Dr. Susie Sinks for their commitment to my dissertation. I appreciate their comments and suggestions. To Dr. McClish, thanks for being my academic advisor and giving me advice on my graduate study. And thank Dr. Pakyz for providing me the opportunity to work as a research assistant at School of Pharmacy for the past two years. I also appreciate the kindness, support, and help from my classmates and the rest of faculty and staff in the department.

Finally, I thank my family, especially my parents and my uncle and aunt, without whose support, encouragement, and love, I would not have finished this thesis.

Contents

List of Tables

List of Figures

Abstract

RESPONSE ADAPTIVE RANDOMIZATION USING SURROGATE AND PRIMARY ENDPOINTS

By Hui Wang

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University.

Virginia Commonwealth University, 2016

Major Director: Nitai Mukhopadhyay, Ph.D., Associate Professor, Department of Biostatistics

In recent years, adaptive designs in clinical trials have been attractive due to their efficiency and flexibility. Response adaptive randomization procedures in phase II or III clinical trials are proposed to appeal ethical concerns by skewing the probability of patient assignments based on the responses obtained thus far, so that more patients will be assigned to a superior treatment group. General response-adaptive randomizations usually assume that the primary endpoint can be obtained quickly after the treatment. However, in real clinical trials, the primary outcome is delayed, making it unusable for adaptation. Therefore, we utilize surrogate and primary endpoints simultaneously to adaptively assign subjects between treatment groups for clinical trials with continuous responses. We explore two types of primary endpoints commonly used in clinical tirials: normally distributed outcome and time-to-event outcome. We establish a connection between the surrogate and primary endpoints

through a Bayesian model, and then update the allocation ratio based on the accumulated data. Through simulation studies, we find that our proposed response adaptive randomization is more effective in assigning patients to better treatments as compared with equal allocation randomization and standard response adaptive randomization which is solely based on the primary endpoint.

Chapter 1

Introduction

1.1 Adaptive Design

Clinical trials are prospective intervention studies with human subjects to investigate experimental drugs, new treatments, medical devices, or clinical procedures, under rigorously specified conditions [Yin, 2013]. Traditionally, clinical trials are designed with fixed sample size and allocation probabilities among treatment group. No changes can be made while the trial goes on and accumulated information becomes available. In 2006, the United States Food and Drug Administration (FDA) released a critical path opportunities list which recommended the creation of innovative and efficient clinical trials that apply the accumulated information in the trial design [Food et al., 2007]. Specidically, the FDA began encouraging the use of adaptive design methods in clinical trials.

Adaptive designs in clinical trials are attractive due to their efficiency and flex-

ibility. However, there is no universal definition of it. In 2006, the pharmaceutical research and Manufacturers of America (PhRMA) defined an adaptive design as a clinical trial design that uses accumulating data to decide on how to modify aspects of the study as it continues, without undermining the validity and integrity of the trial[Gallo et al., 2006]. Based on this definition, adaption is the main feature to improve the design. In 2010, the U.S. Food and Drug Administration (FDA) released a guidance on the regulatory aspects of adaptive designs[Administration et al., 2010]. The FDA explains that there is great interest in the possibility that clinical trials can be designed with adaptive features that may make the studies more efficient (e.g., shorter duration, fewer patients), more likely to demonstrate a treatment effect, or more informative (e.g., broader dose-response findings). In the guidance, the adaptive design clinical study is referred to as a study that includes a prospectivly planned opportunity for modification of one or more specified aspects of the study design and hypotheses based on analysis of data (usually interim data) from subjects in the study. Based on adaptations employed in general, commonly considered adaptive design methods in clinical trials include adaptive randomization, group sequential methods, sample size re-estimation, adaptive dose finding, adaptive treatment-switching, etc[Chow et al., 2008]. Another way of classifying adaptive design clinical trials is by categorizing them based on four different rules: Allocation rule, sampling rule, stopping rule, and decision rule[Mahajan and Gupta, 2010]. Allocation rules define how subjects are allocated to different arms in a trial, which includes response-adaptive randomization and covariate adaptive allocation. Sampling rules define how many subjects will be sampled at the next stage, which

includes sample size re-estimation and drop-the-loser designs. Stopping rules define when to stop the trial, which includes group sequential design and adaptive treatment-switching design. Decision rules refer to changes not covered under the other three categories, which may include hypothesis-adaptive design and change the primary end-point or statistical method or patient population design.

1.2 Response Adaptive Randomization

In clinical trials, patients are assigned to different groups to receive different treatments. The process of assigning patients to different groups by chance is called randomization. The primary goal of randomization is to prevent bias in allocating subjects to treatment groups, thereby obtaining a credible and unbiased result. Traditionally, equal randomization or randomization with a fixed ratio (e.g. $1:3$) among the groups are commonly used in clinical trials. The main feature of traditional randomization is that the probability of assigning patient to each treatment group is fixed and pre-determined. Even through when the trial proceeds, one may find that the treatment group performs better than the control group, the allocation probability can never be changed. However, there is one ethical concern that more patients should be assigned to the better treatment group if evidence of superiority exists. Recently, the response adaptive randomization or RAR becomes popular since it can change the allocation probabilities when the trial goes on.

Response-adaptive randomization procedures in clinical trials are proposed to appeal the ethical concerns by skewing the probability of patient assignments based

on the responses obtained thus far, so that more patients will be assigned to a superior treatment group[Zhang and Rosenberger, 2006]. The preliminary ideas of response adaptive design was proposed by Thompson (1933) and Robins (1952), and have been further developed by other researchers. Most of the available works in adaptive randomization designs have been focused on binary outcomes[Biswas and Bhattacharya, 2012]. Examples include the play-the-winner rule (Zelen, 1969), the randomized play-the-winner rule (Wei and Durham, 1978), the success-driven design (Durham, Flournoy and Li, 1998), the drop-the-loser design (Ivanova, 2003), and the generalized drop-the-loser design (Zhang, Chan and Cheung, 2007), etc. However, many real clinical trials need to deal with continuous outcomes. For example, Wilson et al. (1998) used office-recorded diastolic blood pressure reduction, which may be considered to have an approximate normal distribution, as the primary outcome to evaluate the antihypertensive efficacy of losartan and amlodipine[Wilson et al., 1998]. Fu et al. (1998) considered wound healing time, which is a time to event outcome, as the primary outcome to evaluate the efficacy of topical recombinant bovine basic fibroblast growth factor for second-degree burns[Fu et al., 1998].

In 1993, Rosenberger introduced a reasonable allocation design for the case of general (not necessarily dichotomous) responses[Rosenberger et al., 1993]. He used the idea of treatment effect mapping, in which the allocation probabilities are functions of the current estimate of treatment effect. This method can be used for clinical trials with continuous responses to skew the allocation to the better treatment. Later on, some other response-adaptive designs have been developed for continuous responses in clinical trials, such as the continuous drop-the-loser rule, Wilcoxon-

Mann-Whitney-type adaptive design, doubly adaptive biased coin designs (DBCD), Kernel-based allocation designs, etc. Except for DBCD, all the other designs mentioned above are not based on any optimal consideration. For those designs which are not based on any optimality criteria, they may have high variability which may lead to significant loss in power[Zhang and Rosenberger, 2006]. When a smaller response is desirable, Zhang and Rosenberger (2006) suggested that the DBCD procedure with optimal allocation can reduce total expected response and simultaneously maintain power, and should be the first choice for response-adaptive randomization designs with continuous outcome. So our response-adaptive design will be an extension of the doubly adaptive biased coin design procedure with optimal allocation. More details of this procedure will be discussed in Chapter 2.

1.3 Motivation

The general response-adaptive randomization mentioned above assumes that the primary endpoint can be obtained quickly after the treatment. However, in real clinical trials, and not only for survival outcomes, one may need to take a relatively long time to observe the primary endpoints. Some studies have been done on delayed primary endpoints and their effect on response-adaptive randomization in the literature. Zhang and Rosenberger (2006) explored the effect of delayed responses on response-adaptive randomization for continuous outcomes based on some simulation studies[Zhang and Rosenberger, 2006]. They found that moderate delays in responses have little effect on the power and asymptotic properties of the DBCD procedure.

Later on, Hu et al. studied the effects of delayed responses in DBCD mathematically[Hu et al., 2008]. They found that the asymptotic properties of the allocation proportions are unaffected by staggered entry and delayed responses in reasonable probability models. Even though moderate delays in responses have no effect on the asymptotic properties of randomization procedure under certain delay mechanisms, the allocation rate through the trial is directly affected and there is a higher risk of assigning more patients to the inferior treatment[Sinks, 2013].

We are motivated to propose a new response-adaptive randomization method that incorporates one or more surrogate endpoints, that are correlated with the primary endpoint. In clinical trials, a biomarker is defined as a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention[Group, 2001]. A surrogate endpoint is a biomarker that is intended to substitute for a clinical endpoint. Even though a surrogate endpoints can be selected so that it is measured earlier than the primary endpoint, the incorporation of that in response-adaptive randomization has not been fully explored. Huang et al. (2009) perhaps are the first who use short-term response information to facilitate adaptive randomization in clinical trials. They proposed a new design for survival trials that connects short-term response with long-term survival to 'speed up' the adaptation of the randomization procedure[Huang et al., 2009]. In their study, they assume that conditional on the categorical short-term responses, the long-term response follows an exponential distribution. However, they didn't optimize the allocation design in their study. Sinks (2013) proposed a bivariate response adaptive design for binary

primary endpoint where a binary auxiliary endpoint was used to assist the adaptation with the primary endpoint [Sinks, 2013]. She found that the bivariate adaptive design was more effective in assigning patients to better treatments as compared with univariate optimal and balanced designs. Rencently, Nowacki et al. (2015) proposed a surrogate-primary (S-P) replacement algorithm where a patent's surrogate outcome is used in the response-adaptive randomization only until their primary outcome becomes available to replace it[Nowacki et al., 2015]. They showed that the S-P replacement algorithm performs better than the standard approach by reducing the probability variability and increasing convergence of the treatment allocation ratio toward its target.

In my dissertation study, we will focus on clinical trials with continuous primary outcomes. We assume that there is a delay in the primary endpoint, but the surrogate endpoint which is associated with the primary endpoint can be observed immediately. We will propose a new response-adaptive randomization method that will utilize surrogate and primary endpoints at the same time for clinical trials with continuous primary responses. The connection of surrogate and primary endpoints is established through a Bayesian model. We will draw inferences about parameters of interest through a Markov Chain Monte Carlo (MCMC) simulation. Then these parameters will be used to calculate the desired target allocation to estimate the optimal proportion for allocating subjects between treatments. Finally, we will use these sequentially estimated proportions based on DBCD rule to skew the allocation. Details will be discussed in later chapters.

1.4 Organization of this thesis

In Chapter 2, we will review some current statistical approaches of response adaptive randomization for clinical trials with continuous outcomes. These approaches are solely based on the primary endpoint, and always assume that the primary endpoint can be observed immediately after treatment. Even through researchers have shown that for continuous response, the delayed response has no effect on the asymptotic properties of the randomization procedure under very general conditions, we still believe that the delay in the primary response will influence the performance of the standard response adaptive randomization, where only the accumulated information from primary endpoint is considered during the randomization procedure. To illustrate this issue, we will conduct some simulation studies with different proportion of delays in the primary endpoint to show the drawback of the standard response adaptive randomization.

In Chapter 3, we will consider a clinical trial where both surrogate and primary endpoints have a normal distribution and propose a new response adaptive randomization procedure which will extend the standard adaptive randomization design for normally distributed continuous outcomes to simultaneously account for the surrogate endpoint. We will refer to the allocation procedure proposed by Zhang and Rosenberger[Zhang and Rosenberger, 2006] as the univariate optimal adaptive design for continuous outcomes. Under this procedure, the optimal allocation ratio is obtained by minimizing the total expected response of patients if we assume a lower response is desirable. To implement the surrogate endpoint in the adaptation process, we assume that the surrogate and primary endpoints follow a bivariate nor-

mal distribution. Thus given the surrogate endpoint, the conditional distribution of the primary endpoint will also be normal. We will start with the conditional model of primary endpoint given surrogate to perform the adaptive randomization procedure. Conjugate priors will be given to the parameters in the conditional model. The allocation ratio for each subject will be calculated based on the posterior mean of parameters in the conditional model. A simulation study will be performed to compare the performance, specifically in power and treatment assignment skewing to better treatment, of simple randomization, univariate optimal adaptive randomization, and bivariate optimal adaptive randomization.

In Chapter 4, we will focus on clinical trials where the primary endpoint is a time-to-event outcome. Specifically, we assume that the primary endpoint follows an exponential distribution and there is a categorical surrogate endpoint that can be obtained immediately after the treatment. Then we will propose a new randomization algorithm which will extend the standard adaptive randomization for survival trials to simultaneously account for a binary surrogate endpoint. We will refer to the allocation procedure proposed by Zhang and Rosenberger as the univeriate optimal adaptive design for survival outcomes [Zhang and Rosenberger, 2007]. Under this procedure, the optimal allocation proportion is obtained by minimizing the total expected hazard. To implement the surrogate endpoint in the randomization procedure, we assume that the surrogate endpoint follows a multinomial distribution and given the surrogate endpoint, the primary endpoint follows a mixture of exponential distribution. We will estimate the parameters of interest through a Bayesian model, and then apply these estimates to the optimal allocation function to skew the al-

location probability. The performance of the proposed algorithm will be evaluated through a series of simulations.

In Chapter 5, we will discuss the two randomization procedures we proposed in the previous two chapters, and some future works that can be done.

Chapter 2

Current statistical approaches of response-adaptive randomization for continuous outcomes

Response-adaptive designs can be classified into two categories. The first is the target-driven response-adaptive design that is based on an optimal allocation target, where a specific criterion is optimized based on a population response model. The second class is the design-driven response-adaptive randomization, where allocation rules are established with an intuitive motivation, but not optimal in a formal sense[Rosenberger and Lachin, 2004]. In this dissertation study, the proposed randomization procedure in chapter 3 and chapter 4 will be based on the first category, which is target-driven response adaptive randomization. So in this chapter, we are only going to review some commonly used optimal response-adaptive designs in

clinical trials with continuous outcome.

2.1 Doubly adaptive biased coin designs

Doubly adaptive biased coin design (DBCD) is a family of response-adpative procedures that can be used to target a desired allocation proportion ρ. It was first proposed by Eisele in 1994 and then sequentially modified by some other researchers [Eisele and Woodroofe, 1995].

Eisele (1994) and Woodroffe (1995) proposed a DBCD procedure for two treatment groups to target any desired allocation proportion ρ to treatment A. They defined a function $g(x, y)$ on $[0, 1]^2 \times [0, 1]$ which needs to satisfy the following conditions:

(i) g is jointly continuous;

(ii) $g(x, x) = x$,

(iii) $g(x, y)$ is strictly decreasing in x and strictly increasing in y on $[0, 1]^2$, and

(iv) g has bounded derivatives in both arguments.

Suppose j subjects have been randomized to the two treatment groups, then the DBCD procedure will allocate the $(j+1)$th subject to treatment A with probability $g(n_{Aj}/j, \hat{\rho}_j)$, where n_{Aj} is the number of subjects assigned to treatment A so far, and $\hat{\rho}_j$ is the estimated target allocation proportion based on the first j subjects. Thus, the double adaptive biased coin design (DBCD) depends on both the current observed allocation proportion and the estimate of target allocation proportion.

The key component of this procedure is the choice of an appropriate allocation

function $g(x, y)$. Hu and Zhang (2004) proposed the following family of allocation function:

$$g^{(\alpha)}(x,y) = \begin{cases} 1 & \text{if } x = 0 \\ \frac{y(\frac{y}{x})^{\alpha}}{y(\frac{y}{x})^{\alpha}+(1-y)(\frac{1-y}{1-x})^{\alpha}} & \text{if } 0 < x < 1 \\ 0 & \text{if } x = 1 \end{cases}$$

where $\alpha \geq 0$ and controls the degree of randomness of the procedure[Hu and Zhang, 2004]. Different choices of α lead to different allocation procedures. For example, if $\alpha = 0$, then $g^{(\alpha)}(x, y) = y$ and we will have the Sequential Maximum Likelihood Estimation, where at each stage the target allocation proportion is estimated, preferably by the maximum likelihood method, and the next incoming subject will be assigned to treatment A with this probability. In general, a DBCD procedure with a large value of α will provide a smaller variance but will decrease the degree of randomness[Biswas and Bhattacharya, 2012]. Therefore, α should be chosen to reflect the trade-off between the degree of randomness and the variation [Atkinson and Biswas, 2013]. Suppose N_A is the number of subjects assigned to treatment A and n is the total sample size. Hu and Zhang (2004) showed that under mild conditions, both N_A/n and $\hat{\rho}$ converge to ρ almost surely, and have an asymptotic bivariate normal distribution.

2.2 Optimal response-adaptive designs

As we mentioned before in section 2.1, the double adaptive biased coin design (DBCD) depends on the target allocation proportion. For ethical consideration in

response-adaptive randomization in clinical trials, it is desired to target an allocation proportion which is optimal in some sense[Zhang and Rosenberger, 2006].

2.2.1 General approach

To develop an optimal allocation design, we need to define a clinically relevant criterion and then try to optimize it. In 2007, Biswas et al. proposed a general approach to obtain the optimal target allocation for clinical trials with continuous outcomes[Biswas et al., 2007]. Consider a clinical trial with two treatment groups and let k be the treatment indicator, where $k = 1, 2$. Let X_k be the primary outcome which is assumed to have a continuous distribution with mean μ_k and finite variance σ_k^2, $k = 1, 2$. Let n_1 and n_2 be the sample sizes for the two treatments in a non-randomized manner. Then for the general approach, one may choose to minimize

$$n_1\Psi_1 + n_2\Psi_2$$

subject to the restriction

$$\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} = K$$

where Ψ_k is a positive function of (μ_k, σ_k), which will be different for different goal of the trial, and K is a constant. Solving this optimization problem, we can obtain the optimal allocation proportion to treatment 1 as below:

$$\rho = \frac{\sigma_1\sqrt{\Psi_2}}{\sigma_1\sqrt{\Psi_2} + \sigma_2\sqrt{\Psi_1}}$$

Here, ρ depends on unknown parameters and these unknown parameters can be estimated through either frequentist or Bayesian method.

2.2.2 Allocations for normally distributed responses

Now consider the situation where the primary endpoint has normal distribution. We assume that the primary outcomes for each treatment group are normally distributed with $X_1 \sim \mathrm{N}(\mu_1, \sigma_1^2)$ and $X_2 \sim \mathrm{N}(\mu_2, \sigma_2^2)$, respectively. Without loss of generality, we assume that a small response is desirable. As we mentioned before, different choices of $\Psi_1 = g(\mu_1, \sigma_1)$ and $\Psi_2 = g(\mu_2, \sigma_2)$ can be made, and these Ψ_1 and Ψ_2 reflect different goals or targets of the trial. So far, only a few optimal adaptive designs are available for continuous responses with normal distribution. In this section, we will summarize some of these available allocation proportions that are commonly discussed in the literature.

If $\Psi_k = 1$ for each of the two treatment groups, then the optimization problem is to minimize the total sample size for a fixed variance of the estimated treatment difference[Biswas and Bhattacharya, 2012]. The solution for this optimization problem is the well-known Neyman allocation which can be expressed as

$$\rho = \frac{\sigma_1}{\sigma_1 + \sigma_2} \tag{2.1}$$

where ρ is the optimal allocation proportion to treatment 1.

If we let $\Psi_k = \mu_k$, for $k = 1, 2$, then the optimization problem is equivalent to minimize the total expected response from all patients. This optimization problem

was proposed by Zhang and Rosenberger in 2006 [Zhang and Rosenberger, 2006]. Define $r = \sigma_1\sqrt{\mu_2}/\sigma_2\sqrt{\mu_1}$ and

$$s = \begin{cases} 1 & \text{if } (\ \mu_1 < \mu_2 \text{ and } r > 1) \text{ or } (\mu_1 > \mu_2 \text{ and } r < 1), \\ 0 & \text{otherwise.} \end{cases}$$

Then the optimal allocation proportion to treatment 1 for this problem will be

$$\rho = \begin{cases} \frac{\sigma_1\sqrt{\mu_2}}{\sigma_1\sqrt{\mu_2}+\sigma_2\sqrt{\mu_1}} & \text{if } s = 1, \\ \frac{1}{2} & \text{otherwise.} \end{cases} \tag{2.2}$$

Furthermore, another choice of Φ_k could be $\Psi_k = \Phi(\frac{\mu_k - c}{\sigma_k})$ where c is a predefined threshold constant and $\Phi(\cdot)$ is the cumulative density function of standard normal distribution. This optimization problem was proposed by Biswas and Mandal in 2004 and the aim is to minimize the total number of patients with response greater than some threshold constant[Biswas and Mandal, 2004]. Solving this optimization problem, we can obtain the optimal allocation proportion to treatment 1 as below:

$$\rho = \frac{\sqrt{\Phi(\frac{\mu_2 - c}{\sigma_2})}\sigma_1}{\sqrt{\Phi(\frac{\mu_2 - c}{\sigma_2})}\sigma_1 + \sqrt{\Phi(\frac{\mu_1 - c}{\sigma_1})}\sigma_2}$$

Zhang and Rosenberger (2006) compared the performance of the above three randomization procedures through a simulation study, and they found that the DBCD procedure with their optimal allocation proportion can reduce the total expected response and simultaneously maintain the power, and thus should be the first choice

for response adaptive randomization with continuous outcomes [Zhang and Rosenberger, 2006]. Thus in chapter 3, where the primary endpoint is assumed to be normally distributed, we will extend the optimal allocation procedure proposed by Zhang and Rosenberger in equation 3.1 to simultaneously account for the information from surrogate endpoint.

2.2.3 Allocations for survival responses

For clinical trials with survival or time-to-event outcomes, inherent delay is common. Zhang and Rosenberger (2007) developed a response-adaptive randomization procedure for survival outcomes with censoring and delay by using a parametric approach that involves a target optimal allocation and a randomization procedure with low variability[Zhang and Rosenberger, 2007]. Consider a clinical trial with two treatment groups, and k is the treatment indicator, where $k = 1, 2$. Let n_k be the number of subjects assigned to treatment k. Now we assume that for the ith subject in treatment k, the survival time T_{ik} follows an exponential distribution with mean θ_k. Furthermore, let C_i be the censoring time and is assumed to be independent of T_{ik}. Then for each subject i, one observes a pair of random variables (t_{ik}, δ_{ik}), where $t_{ik} = \min(T_{ik}, C_i)$ is the observed time and δ_{ik} is an indicator of event ($\delta_{ik} = 1$ if the ith subject has an event; and $\delta_{ik} = 0$ if the ith subject is censored). Assume $\epsilon_k = E(\delta_{ik})$ is the same for subjects in the same treatment group. Under the above assumption, Zhang and Rosenberger (2007) proposed an optimal allocation proportion where the target is to minimize the total expected hazard. The optimization

problem can be stated below:

$$\begin{cases} \min\limits_{n_A/n_B} & n_1\theta_1^{-1} + n_2\theta_2^{-1} \\ \text{s.t.} & \theta_1^2/n_1\epsilon_1 + \theta_2^2/n_2\epsilon_2 = K \end{cases}$$

In this case, $\Psi_k = \theta_k^{-1}$ and $\sigma_k^2 = \frac{\theta_k^2}{\epsilon_k}$. And the corresponding optimal allocation proportion to treatment 1 is

$$\rho = \frac{\sqrt{\theta_1^3\epsilon_2}}{\sqrt{\theta_1^3\epsilon_2} + \sqrt{\theta_2^3\epsilon_1}}$$

Another choice could be $\Psi_k = 1$ for $k = 1, 2$, and the aim is to minimize the total number of subjects in the trial. Solving this problem will obtain the Neyman allocation, where the optimal allocation proportion to treatment 1 is

$$\rho = \frac{\theta_1\sqrt{\epsilon_2}}{\theta_1\sqrt{\epsilon_2} + \theta_2\sqrt{\epsilon_1}}$$

2.3 Delayed responses

An important assumption for the double adaptive biased coin design was that the primary outcome can be observed immediately after the treatment. However, it is common that responses may not be available before the randomization of next subject and will be available after a period of time. Hu et al.(2008) examined the effect of delayed responses on DBCD procedures and derived some of their asymptotic properties[Hu et al., 2008]. Suppose that the subjects arrive sequentially in a clini-

cal study, and the subjects delay time is not unreasonably large compared to their arrival time, then they found out that under widely satisfied conditions for the delay mechanism, the asymptotic properties of DBCD procedure remain insensitive to a stochastic delay in updating the sequential estimator of the unknown parameters. However, in real clinical trials, the delay in the primary endpoint will influence the implementation of adaptive randomization. For example, when no primary outcome becomes available for the next patient, then no information can be used to skew the allocation proportion, and thus a simple randomization (e.g. equal allocation) needs to be used for this patient. To illustrate this issue, we will conduct some simulations.

Consider a clinical trial with two treatment groups, and the primary outcome is continuous. To be specific, suppose the primary outcome is normally distributed, such that $X_A \sim \mathrm{N}(\mu_A, \sigma_A^2), X_B \sim \mathrm{N}(\mu_B, \sigma_B^2)$, respectively. Furthermore, suppose there is a constant enrollment rate. Therefore, there is a fixed delay time in the primary endpoint. For example, to observe the primary outcome from the first patient, 20 more patients enroll in the trial. Which means, the response of the first patient will be available when the 21st patient enrolls in the trial, and the response of the second patient will be available when the twenty second patient enrolls in the trial, etc. Figure 2.1 shows the effect of delay in the primary responses for the standard response adaptive randomization (only consider the primary endpoint).

As we can see from the plot, if there is no delay in the primary endpoint (all the primary outcomes are available), then the proportion pf patients assigned to treatment A is equal to the target allocation proportion. However, as the delay in the primary endpoint increases (the proportion of primary outcomes available

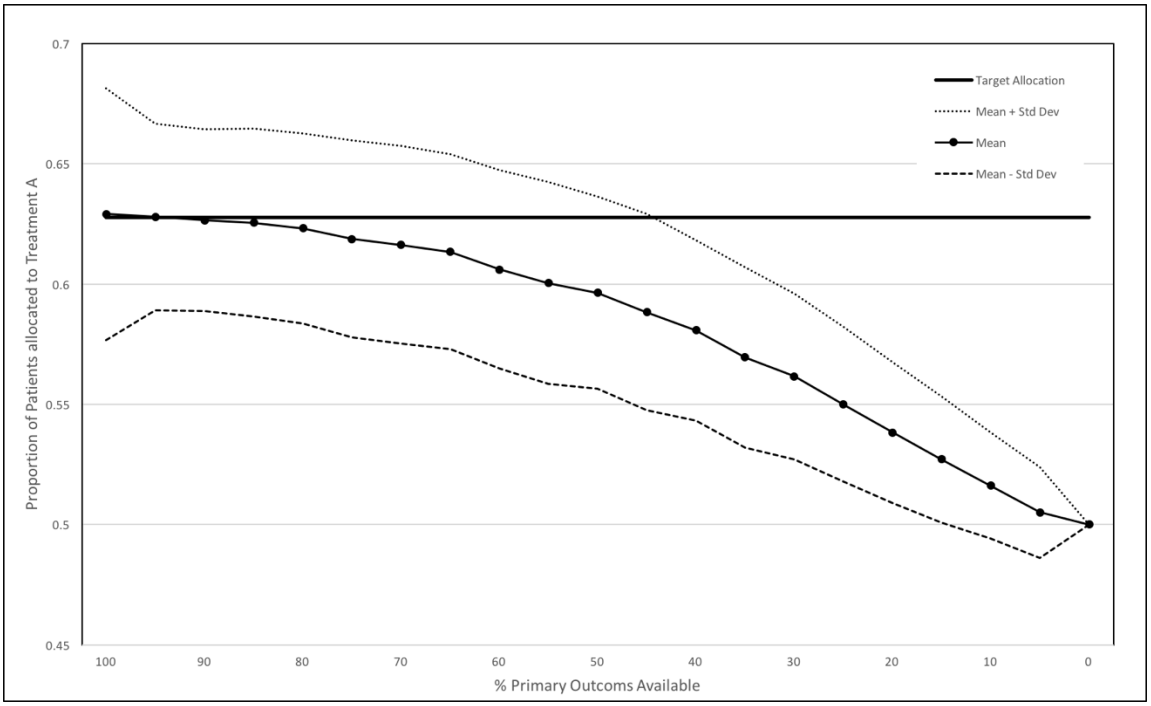


Figure 2.1: The effect of delay on treatment allocation when using standard response adaptive randomization.

decreases), the proportion of patients assigned to treatment A decreases. And when there is no primary outcome available, the proportion equals to 0.5, which is the same as the traditional equal randomization. This means that as the delay time increases, the difference between the target allocation proportion and the observed allocation proportion becomes larger, and thus the benefit of assigning more patients to a better treatment group disappears. Moreover, we can also see from the plot that the variability of the allocation proportion to treatment A is getting larger when more primary outcomes become available. That is because the allocation probability changes whenever a new patient enrolls.

The standard response adaptive randomization is influenced by the delay in the

primary endpoint. The benefit of assigning more patients to a better treatment group disappears when there is a large delay time in the primary endpoint. Thus, we are motivated to propose a new algorithm to simultaneously account for the information from the surrogate endpoint, such that the benefit of assigning more patients to a better treatment group will not disappear even when there is a large delay time in the primary response.

Chapter 3

Bivariate Response Adaptive Design for Continuous Outcomes

In this chapter, we will propose a new response adaptive randomization for clinical trials with normally distributed primary endpoint. In section 3.1, we will have a brief review of the standard response adaptive randomization for clinical trials with normally distributed primary endpoint. Then we will extend the standard adaptive randomization procedure to simultaneously account for the information from the surrogated endpoint by assuming a correlation between the surrogate and primary endpoints in section 3.2. Finally in section 3.3, we will compare the proposed algorithm with the standard response adaptive randomization and traditional equal allocation procedure through some simulation studies.

3.1 Response adaptive randomization using primary endpoint only

In the classical response adaptive procedure, consider a clinical trial with only two treatment groups: Treatment A and Treatment B, and we want to compare the difference between the two treatments with normally distributed responses, such that $T_A \sim N(\mu_{T,A}, \sigma^2_{t,A})$ and $T_B \sim N(\mu_{T,B}, \sigma^2_{t,B})$, respectively. Suppose a smaller response is desirable and we will consider the randomization procedure proposed by Zhang and Rosenberger (2006) as the standard response adaptive randomization (RAR), where the optimization problem is:

$$\begin{cases} \min\limits_{n_A/n_B} & \mu_{T,A} n_A + \mu_{T,B} n_B \\ \text{s.t.} & \frac{\sigma^2_{t,A}}{n_A} + \frac{\sigma^2_{t,B}}{n_B} = K \end{cases}$$

n_A and n_B refer to the cumulative number of subjects assigned to treatment A and treatment B respectively, and K is some constant. Solving this problem yields the optimal allocation proportion to treatment A as

$$\rho = \frac{\sqrt{\mu_{T,B}}\sigma_{t,A}}{\sqrt{\mu_{T,B}}\sigma_{t,A} + \sqrt{\mu_{T,A}}\sigma_{t,B}}$$

Then they defined $r = \sigma_1\sqrt{\mu_2}/\sigma_2\sqrt{\mu_1}$ and

$$s = \begin{cases} 1 & \text{if } (\ \mu_1 < \mu_2 \text{ and } r > 1) \text{ or } (\mu_1 > \mu_2 \text{ and } r < 1), \\ 0 & \text{otherwise.} \end{cases}$$

and modified the optimal allocation proportion to treatment A as

$$\rho = \begin{cases} \frac{\sigma_1\sqrt{\mu_2}}{\sigma_1\sqrt{\mu_2}+\sigma_2\sqrt{\mu_1}} & \text{if } s = 1, \\ \frac{1}{2} & \text{otherwise.} \end{cases} \tag{3.1}$$

As we mentioned before in Chapter 2, the Doubly Adaptive Biased Coin Design (DBCD) procedure can be used to target any allocation proportion ρ. Throughout this paper, we will use the DBCD procedure proposed by Hu and Zhang (2004)

$$g^{(\alpha)}(x,y) = \begin{cases} 1 & \text{if } x = 0 \\ \frac{y(\frac{y}{x})^\alpha}{y(\frac{y}{x})^\alpha+(1-y)(\frac{1-y}{1-x})^\alpha} & \text{if } 0 < x < 1 \\ 0 & \text{if } x = 1 \end{cases} \tag{3.2}$$

to skew the allocation at each randomization stage to target our optimal allocation proportion. Here, α is a nonnegative number that controls the randomness of the procedure. Then the probability of assign the $(j+1)th$ subject to treatment A is $g(n_{Aj}/j, \hat{\rho}_j)$, where n_{Aj}/j is the observed proportion of subjects assigned to treatment A so far, and $\hat{\rho}_j$ is the estimate of ρ in equation (3.1) after the j subjects. Hu and Zhang (2004) have shown that both N_A/n and $\hat{\rho}$ approach to the target allocation proportion ρ as n increases.

For the above standard response adaptive randomization, the randomization procedure is solely based on the information from the primary endpoints. In that case, this procedure will perform better when the primary endpoint can be observed immediately relative to the enrollment period. In practice, however, one may need a

relatively long time to obtain the primary outcome. When there are delayed primary responses, less information will be available to skew the allocation.

3.2 Proposed method

Many clinical trials have surrogate endpoints that can by observed sooner than the primary endpoint. These surrogate endpoints will provide additional information about the primary endpoints which can be used in the randomization procedure. In this section, we will propose a new response adaptive randomization design which will extend the standard RAR to simultaneously account for the information from the surrogate endpoint.

3.2.1 Likelihood distribution

Suppose S_i and T_i are the surrogate and primary endpoints for the ith subject, respectively. We assume that both the surrogate and primary endpoints have a normal distribution. Specifically, we assume that S_i and T_i follow a bivariate normal distribution with mean vector μ and variance-covariance matrix Σ_L. That is

$$\begin{pmatrix} S_i \\ T_i \end{pmatrix} \sim N_2 \left(\begin{pmatrix} \mu_S \\ \mu_T \end{pmatrix}, \begin{pmatrix} \sigma_s^2 & \sigma_{st} \\ \sigma_{st} & \sigma_t^2 \end{pmatrix} \right) \tag{3.3}$$

where

$$\mu = \begin{pmatrix} \mu_S \\ \mu_T \end{pmatrix} \quad \text{and} \quad \Sigma_L = \begin{pmatrix} \sigma_s^2 & \sigma_{st} \\ \sigma_{st} & \sigma_t^2 \end{pmatrix}.$$

Based on the properties of the bivariate normal distribution, the conditional distribution of primary endpoint given surrogate endpoint for subject i is also a normal distribution:

$$T_i|S_i = s_i \sim N(\mu_{t_i|s_i}, \sigma^2_{t|s}) \tag{3.4}$$

where the conditional mean and variance can be expressed as:

$$\mu_{t_i|s_i} = \mu_T + \sigma_{st}\sigma_s^2(s_i - \mu_S)$$
$$\sigma^2_{t|s} = \sigma_t^2 - \sigma^2_{st}\sigma_s^{-2}$$

The contribution to the likelihood from subject i is the joint distribution of surrogate and primary endpoint $f(s_i, t_i)$, which can be expressed as a product of two parts: the marginal distribution of the surrogate endpoint $f(s_i)$, and the conditional distribution of primary endpoint given the surrogate endpoint $f(t_i|s_i)$. Let $D = (S_1, \cdots, S_n, T_1, \cdots, T_n)$ be the data that contains the surrogate and primary endpoints, and $\Theta_L = (\mu_S, \mu_T, \sigma_s^2, \sigma_t^2, \sigma_{st})$ be the vector of unknown parameters, then the likelihood for all the parameters in the model will be:

$$\begin{aligned} L_n(\Theta_L|D) &= \prod_{i=1}^{n} f(s_i, t_i) = \prod_{i=1}^{n} f(t_i|s_i)f(s_i) \\ &= \prod_{i=1}^{n} \left[\frac{1}{2\pi\sqrt{\sigma_s^2}} \exp\left\{ -\frac{(s_i - \mu_S)^2}{2\sigma_s^2} \right\} \times \frac{1}{2\pi\sqrt{\sigma^2_{s|t}}} \exp\left\{ -\frac{(t_i - \mu_{t_i|s_i})^2}{2\sigma^2_{t|s}} \right\} \right] \end{aligned} \tag{3.5}$$

3.2.2 Prior distribution

In this section, we will explore the use of conjugate priors for the purpose of obtaining the posterior distribution of conditional mean and variance.

Normal-Inverse-Wishart Distribution

In probability theory and statistics, the Normal-inverse-Wishart(NIW) distribution is a multivariate four-parameter family of continuous probability distributions. It is the conjugate prior of a multivariate normal distribution with unknown mean and unknown variance-covariance matrix[Murphy, 2007]. In our study, we assume that the joint prior distribution of mean vector μ and variance-covariance matrix Σ_L is a Normal-inverse-Wishart distribution, such that $(\mu, \Sigma_L) \sim NIW(\mu_0, \kappa_0, R, \upsilon)$. Under the above prior assumption, the prior distribution of mean μ is dependent on the prior of variance-covariance matrix Σ_L. Specifically, this prior distribution can be expressed as below:

$$\mu|\Sigma_L \sim N(\mu_0, \Sigma_L/\kappa_0)$$
$$\Sigma_L \sim IW(R, \upsilon)$$

where μ_0 is a parameter vector, κ_0 is the sample size that the prior belief about μ is equivalent to, R is a parameter matrix, and υ is the degrees of freedom.

Then the prior distribution of (μ, Σ_L) can be expressed as

$$\begin{aligned} p(\mu, \Sigma_L) &= p(\mu|\Sigma_L)p(\Sigma_L) = N_2(\mu_0, \Sigma_L/\kappa_0)IW(R, \upsilon) \\ &= \frac{|R|^{\frac{\upsilon}{2}}}{2^{\frac{\upsilon p}{2}}\Gamma(\frac{\upsilon}{2})}\left(\frac{2\pi}{\kappa_0}\right)^{-\frac{p}{2}}|\Sigma_L|^{-\frac{\upsilon+p}{2}-1}\exp\left\{-\frac{1}{2}\operatorname{tr}(R\Sigma_L^{-1}) - \frac{\kappa_0}{2}(\mu-\mu_0)'\Sigma_L^{-1}(\mu-\mu_0)\right\} \\ &\propto |\Sigma_L|^{-\frac{\upsilon+p}{2}-1}\exp\left\{-\frac{1}{2}\operatorname{tr}(R\Sigma_L^{-1}) - \frac{\kappa_0}{2}(\mu-\mu_0)'\Sigma_L^{-1}(\mu-\mu_0)\right\} \end{aligned} \tag{3.6}$$

where p is the rank of Σ_L and $p = 2$ in this case.

Partitioning the Inverse-Wishart Distribution

As we mentioned in the previous section, the likelihood function is expressed as a product of the marginal density of surrogate endpoint and the conditional density of the primary endpoint given the surrogate endpoint. In consequence, the prior distribution on Σ_L needs to be partitioned, since the prior distributions should be placed on parameters σ_s^2 and $\sigma_{t|s}^2$[Shi, 2007]. The partitioning can be achieved by applying the normal theory laid out by Dreze and Richard[Dreze and Richard, 1983]. Suppose the $p \times p$ random matrix $\Sigma \in \mathcal{C}^p$ has an Inverse-Wishart distribution, where $\mathcal{C}^p$ denotes the set of $p \times p$ symmetric positive definite matrices. Specifically, suppose $\Sigma \sim IW(R, \upsilon)$, where R is a $p \times p$ matrix and υ is the degree of freedom. The first moment of this density function is

$$E(\Sigma) = \frac{1}{\upsilon - p - 1}R \qquad (\upsilon > p + 1).$$

If both Σ and R can be partitioned as

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, \qquad R = \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix}$$

where Σ_{11} and R_{11} are $p_1 \times p_1$, Σ_{22} and R_{22} are $p_2 \times p_2$, Σ_{12} and R_{12} are $p_1 \times p_2$, and Σ_{21} and R_{21} are $p_2 \times p_1$ matrices. Furthermore, if we define

$$\Sigma_{22.1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \qquad \Omega = \Sigma_{11}^{-1}\Sigma_{12}$$
$$R_{22.1} = R_{22} - R_{21}R_{11}^{-1}R_{12} \qquad \tilde{\Omega} = R_{11}^{-1}R_{12}$$

Then we will have

$$p(\Sigma_{11}, \Omega, \Sigma_{22.1}) = p(\Sigma_{11})p(\Omega|\Sigma_{22.1})p(\Sigma_{22.1}) \tag{3.7}$$

with

$$\Sigma_{11} \sim IW_{p_1 \times p_1}(R_{11}, \upsilon - p_2) \tag{3.8a}$$
$$\Omega|\Sigma_{22.1} \sim MN_{p_1 \times p_2}(\tilde{\Omega}, \Sigma_{22.1} \otimes R_{11}^{-1}) \tag{3.8b}$$
$$\Sigma_{22.1} \sim IW_{p_2 \times p_2}(R_{22.1}, \upsilon) \tag{3.8c}$$

The $\Omega|\Sigma_{22.1}$ in equation (3.8b) is a random matrix rather that a vector, and MN is the matrix normal distribution which is the generalization of the multivariate normal distribution.

Partitioning the Inverse-Wishart prior on Σ_L

In this chapter, since we only consider one surrogate endpoint, the variance-covariance matrix Σ_L is a 2×2 matrix. The following partitioning on both random matrix Σ_L and parameter matrix R are considered:

$$\Sigma_L = \begin{pmatrix} \sigma_s^2 & \sigma_{st} \\ \sigma_{st} & \sigma_t^2 \end{pmatrix} \qquad R = \begin{pmatrix} r_s^2 & r_{st} \\ r_{st} & r_t^2 \end{pmatrix}$$

where σ_s^2, σ_{st}, σ_t^2, r_s^2, r_{st}, and r_t^2 are all scalars. Therefore, the Inverse-Wishart prior distribution for Σ_L can be partitioned into the following three parts:

$$\sigma_s^2 \sim IW_{1\times 1}(r_s^2, \upsilon - 1) \tag{3.9a}$$

$$\Omega_{st}|\sigma_{t|s}^2 \sim MN_{1\times 1}(\tilde{\Omega}_{st}, \sigma_{t|s}^2 \otimes r_s^{-2}) \tag{3.9b}$$

$$\sigma_{t|s}^2 \sim IW_{1\times 1}(r_{t|s}^2, \upsilon) \tag{3.9c}$$

where

$$\sigma_{t|s}^2 = \sigma_t^2 - \sigma_{st}^2\sigma_s^{-2} \qquad \Omega_{st} = \sigma_s^{-2}\sigma_{st}$$
$$r_{t|s}^2 = r_t^2 - r_{st}^2 r_s^{-2} \qquad \tilde{\Omega}_{st} = r_s^{-2} r_{st}$$

The Inverse-Wishart distributions for σ_s^2 in equation (3.9a) and $\sigma_{t|s}^2$ in equation (3.9c) are one-dimensional. The matrix normal distribution in equation (3.9b) is one-dimensional as well. Since the univariate special case of the Inverse-Wishart distribution is the Inverse-Gamma distribution, and the one-dimensional matrix nor-

mal distribution will reduce to univariate normal distribution, the partition of the Inverse-Wishart prior distribution for Σ_L can be rewritten as

$$\sigma_s^2 \sim IG(\frac{\upsilon - 1}{2}, \frac{r_s^2}{2}) \tag{3.10a}$$

$$\Omega_{st}|\sigma_{t|s}^2 \sim N(\tilde{\Omega}_{st}, \sigma_{t|s}^2 r_s^{-2}) \tag{3.10b}$$

$$\sigma_{t|s}^2 \sim IG(\frac{\upsilon}{2}, \frac{r_{t|s}^2}{2}) \tag{3.10c}$$

3.2.3 Estimation rule for allocation rate

In our proposed algorithm, we will use the conditional mean and conditional variance of primary endpoint given surrogate endpoint for each treatment group to update the allocation proportion for the next subject. To make inferences for the unknown parameters, a Bayesian approach will be introduced. The Bayesian approach can be used to estimate the entire posterior distribution, and there is no requirement for the sample size. The Markov chain Monte Carlo (MCMC) method is a general simulation method for sampling from posterior distributions and computing posterior quantities of interest. It provides a convenient computation approach to fitting a Bayesian model of a bivariate normal distribution. Thus, we will use the MCMC method to fit our bivariate normal distribution model.

Suppose there are two treatment groups: A and B. Let $\mu_A = (\mu_{S,A}, \mu_{T,A})'$ be the mean vector for treatment A, where $\mu_{S,A}$ and $\mu_{T,A}$ are the means of the surrogate and primary endpoint for treatment A, respectively. Similarly, $\mu_B = (\mu_{S,B}, \mu_{T,B})'$ is the mean vector for treatment B. Σ_A is the variance-covariance matrix for treatment

A and Σ_B is for treatment B, where the two variance-covariance matrices can be expressed as below:

$$\Sigma_A = \begin{pmatrix} \sigma^2_{s,A} & \sigma_{st,A} \\ \sigma_{st,A} & \sigma^2_{t,A} \end{pmatrix} \qquad \Sigma_B = \begin{pmatrix} \sigma^2_{s,B} & \sigma_{st,B} \\ \sigma_{st,B} & \sigma^2_{t,B} \end{pmatrix}$$

Suppose that the variance-covariance matrix for each treatment group has a Normal-Inverse-Wishart prior as described in section 3.2.2, such that $(\mu_A, \Sigma_A) \sim NIW(\mu_0, \kappa_0, R, \upsilon)$ and $(\mu_B, \Sigma_B) \sim NIW(\mu_0, \kappa_0, R, \upsilon)$. Furthermore, the variance-covariance matrices Σ_A and Σ_B and the parameter matrix R will be partitioned as described in section 3.2.2, such that:

$$\begin{aligned}
&\sigma^2_{s,A} \sim IG(\frac{\upsilon - 1}{2}, \frac{r^2_s}{2}), && \sigma^2_{s,B} \sim IG(\frac{\upsilon - 1}{2}, \frac{r^2_s}{2}) \\
&\Omega_{st,A}|\sigma^2_{t|s,A} \sim N(\tilde{\Omega}_{st}, \sigma^2_{t|s,A} r_s^{-2}), && \Omega_{st,B}|\sigma^2_{t|s,B} \sim N(\tilde{\Omega}_{st}, \sigma^2_{t|s,B} r_s^{-2}) \\
&\sigma^2_{t|s,A} \sim IG(\frac{\upsilon}{2}, \frac{r^2_{t|s}}{2}), && \sigma^2_{t|s,B} \sim IG(\frac{\upsilon}{2}, \frac{r^2_{t|s}}{2})
\end{aligned}$$

where

$$\Omega_{st,A} = \sigma^{-2}_{s,A}\sigma_{st,A}, \qquad \Omega_{st,B} = \sigma^{-2}_{s,B}\sigma_{st,B}$$

Suppose T_i and S_i are the primary and surrogate endpoints for subject i, respectively. Let Z_i denote the treatment indicator for the ith subject, where $Z_i = 1$ if subject i is in treatment A and $Z_i = 0$ if in treatment B. Before assigning the ith subject, we can get the Bayes estimator of $\mu_A = (\mu_{S,A}, \mu_{T,A})'$, $\mu_B = (\mu_{S,B}, \mu_{T,B})'$, $\sigma^2_{s,A}$, $\sigma^2_{s,B}$, $\Omega_{st,A}$, $\Omega_{st,B}$, $\sigma^2_{t|s,A}$, $\sigma^2_{t|s,B}$, which will be denote as $\tilde{\mu}_A = (\tilde{\mu}_{S,A}, \tilde{\mu}_{T,A})'$,

$\tilde{\mu}_B = (\tilde{\mu}_{S,B}, \tilde{\mu}_{T,B})'$, $\tilde{\sigma}^2_{s,A}$, $\tilde{\sigma}^2_{s,B}$, $\tilde{\Omega}_{st,A}$, $\tilde{\Omega}_{st,B}$, $\tilde{\sigma}^2_{t|s,A}$, $\tilde{\sigma}^2_{t|s,B}$, respectively. Besides, the following information can be obtained from the first $i-1$ subjects:

$$n_{A,i} = \sum_{k}^{i-1} Z_k, \qquad n_{B,i} = \sum_{k}^{i-1} (1 - Z_k)$$
$$S_{A,i} = \sum_{k}^{i-1} S_k Z_k, \qquad S_{B,i} = \sum_{k}^{i-1} S_k (1 - Z_k)$$

The above information will be used in the allocation procedure.

3.2.4 Algorithm of the design

The proposed algorithm for response adaptive randomization for clinical trials with normally distributed primary outcome is described below:

S1. Use the equal allocation randomization procedure for a certain number of subjects at the beginning of the trial. Equal randomization is a useful way to obtain initial parameter estimates that are required in a sequential estimation procedure, such as the doubly adaptive biased coin design (DBCD) procedure. As recommended by Nowachi et al., the number of subjects that are equally allocated to treatment A and treatment B in the beginning are typically chosen as $5\% - 10\%$ of the total sample size[Nowacki et al., 2015]. Suppose $2m_0$ subjects are equally assigned to each treatment groups, where $2m_0$ was chosen as 10% of the total sample size.

S2. Before allocating the ith subject, update the following information based on the first $i-1$ subjects:

$n_{A,i}$: current total number of subjects assigned to treatment A;

$n_{B,i}$: current total number of subjects assigned to treatment B;

$\bar{S}_{A,i} = \frac{1}{n_{A,i}} S_{A,i}$: mean of surrogate response assigned to treatment A so far;

$\bar{S}_{B,i} = \frac{1}{n_{B,i}} S_{B,i}$: mean of surrogate response assigned to treatment B so far;

S3. Calculate the Bayes estimators of conditional mean and variance of primary endpoint given surrogate endpoint for each treatment group through a Bayesian model based on the accumulated data:

$\tilde{\sigma}^2_{t|s,A}$: estimate of conditional variance of primary endpoint given surrogate endpoint for treatment A;

$\tilde{\sigma}^2_{t|s,B}$: estimate of conditional variance of primary endpoint given surrogate endpoint for treatment B;

$\tilde{\mu}_{t|s,A} = \tilde{\mu}_{t,A} + \tilde{\Omega}_{st,A}(\bar{S}_{A,i} - \tilde{\mu}_{s,A})$: estimate of conditional mean of primary endpoint given surrogate endpoint for treatment A;

$\tilde{\mu}_{t|s,B} = \tilde{\mu}_{t,B} + \tilde{\Omega}_{st,B}(\bar{S}_{B,i} - \tilde{\mu}_{s,B})$: estimate of conditional mean of primary endpoint given surrogate endpoint for treatment B;

S4. Calculate the current observed allocation proportion for treatment A and the current target allocation proportion for treatment A:

$r_{A,i} = \frac{n_{A,i}}{n_{A,i}+n_{B,i}}$: current observed proportion of subjects assigned to treatment A;

$\hat{\rho} = \frac{\sqrt{\tilde{\mu}_{t|s,B}}\tilde{\sigma}^2_{t|s,A}}{\sqrt{\tilde{\mu}_{t|s,B}}\tilde{\sigma}^2_{t|s,A}+\sqrt{\tilde{\mu}_{t|s,A}}\tilde{\sigma}^2_{t|s,B}}$: current estimate of target allocation proportion to treatment A.

S5. Apply both the current observed treatment A allocation proportion and the current treatment A target allocation proportion calculated in the previous step to the DBCD procedure to calculate the treatment A allocation probability for the ith subject:

$$Prob(TrtA) = \frac{\hat{\rho}(\frac{\hat{\rho}}{r_{A,i}})^{\alpha}}{\hat{\rho}(\frac{\hat{\rho}}{r_{A,i}})^{\alpha} + (1-\hat{\rho})(\frac{1-\hat{\rho}}{1-r_{A,i}})^{\alpha}}$$

where α is a nonnegative number that reflects the desired degree of randomization.

S6. Randomize the next subject using this treatment A allocation probability $Prob(TrtA)$.

S7. Repeat steps 2 - 6 until reaching the predetermined sample size.

3.3 Simulation study

3.3.1 Simulation targets

In this section, some simulation studies will be conducted to investigate the performance of our proposed response adaptive randomization (RAR) for clinical trials with normally distributed primary endpoint and correlated normally distributed surrogate endpoint under different clinical scenarios. Specifically, our proposed algorithm will be compared with equal randomization and standard RAR proposed by Zhang and Rosenberger in 2006 (when only the primary endpoint was considered in the ran-

domization procedure). There are two criteria that can be used to evaluate the performance of response-adaptive randomization procedure[Zhang and Rosenberger, 2006]. One is the total expected response of the primary endpoint which will represent the ethical constraints. The procedure that gives the smallest total expected response will be the most "ethical" procedure. The other criteria that can be used is the power of the test. In our simulation study, we are interested in: 1) how different correlation between primary and surrogate endpoints affect the simulation results, in terms of power, allocation proportions, number of subjects assigned to each treatment group, and total expected response of the primary endpoint; 2) the effect of delay in the primary endpoint; 3) how different sample sizes affect the simulation results.

3.3.2 Sampling method

To make the three allocation procedures (equal randomization, standard RAR, and proposed RAR) comparable, the total sample size will be calculated based on the two-sided two-sample Welch's t-test (unequal variance t-test) when we wish to have equal sample sizes in each treatment groups. More specifically, the sample size is selected to yield a 80% power at significance level of 0.05. The correlated primary and surrogate endpoints are sampled from a bivariate normal distribution. As recommended by Hu and Zhang,the exponential distributions are used for both the delay times of primary endpoint for the two treatment groups and the subject entry time [Hu et al., 2008]. The Gelman-Rubin diagnostic will be used as a numerical support to monitor the convergence of iterative simulations. This approach is, for each parameter of

interest, to compute the variance of the simulations from each chain, to average these within-chain variances, and compare this to the variances of all the chains mixed together [Gelman et al., 2011]. Basically, Gelman-Rubin measures whether there is a significant difference between the variance within several chains and the variance between several chains by a value that is called "scale reduction factors".

3.3.3 Simulation settings

The correlated surrogate and primary endpoints for each treatment groups were generated from the bivariate normal distribution:

$$\begin{pmatrix} S_A \\ T_A \end{pmatrix} \sim N_2\left(\mu_A, \Sigma_A\right), \qquad \begin{pmatrix} S_B \\ T_B \end{pmatrix} \sim N_2\left(\mu_B, \Sigma_B\right)$$

where

$$\mu_A = \begin{pmatrix} \mu_{s,A} \\ \mu_{t,A} \end{pmatrix}, \qquad \Sigma_A = \begin{pmatrix} \sigma^2_{s,A} & \sigma_{st,A} \\ \sigma_{st,A} & \sigma^2_{t,A} \end{pmatrix}$$

and

$$\mu_B = \begin{pmatrix} \mu_{s,B} \\ \mu_{t,B} \end{pmatrix}, \qquad \Sigma_B = \begin{pmatrix} \sigma^2_{s,B} & \sigma_{st,B} \\ \sigma_{st,B} & \sigma^2_{t,B} \end{pmatrix}$$

The correlation between surrogate and primary endpoint will be denoted as $\rho_{st,A}$ and $\rho_{st,B}$, which can be calculated from the variance-covariance matrix. In our simulation study, suppose there is a positive correlation between the surrogate and primary endpoints, and the correlation for each treatment group is the same, which is $\rho_{st,A} = \rho_{st,B} = \rho_{s,t}$. We are going to consider three different strengths for this

relationship: low correlation($0.3 \leq \rho_{s,t} < 0.5$), moderate correlation($0.5 \leq \rho_{s,t} < 0.7$), and high correlation($0.7 \leq \rho_{s,t} < 0.9$). Table 3.1 below shows the possible values of the parameters that we are going to use to generate the data.

Table 3.1: Parameters in the bivariate normal distribution model

Parameters	Values			
$(\mu_{s,A}, \mu_{s,B})$	(20, 24)			
$(\mu_{t,A}, \mu_{t,B})$	(12.5, 15)	(13, 15)	(13.5, 15)	(14, 15)
$(\sigma_{s,A}, \sigma_{s,B})$	(4, 3)			
$(\sigma_{t,A}, \sigma_{t,B})$	(4, 2.5)	(2.5, 4)		
$\rho_{s,t}$	0.35	0.6	0.85	

As mentioned in section 3.3.2 , the exponential distributions were used for the delay times and subject entry times. Suppose the mean parameters of the delay times of primary endpoint for treatment A and B are λ_1 and λ_2, respectively, and the mean parameter for the subject entry times for both treatment groups is λ_3. We are going to consider two different delay scenarios. The first one corresponds to a case where there are similar but moderate delay times for the primary responses of the two treatment groups, such that $(\lambda_1, \lambda_2, \lambda_3) = (1, 1, 1)$. And then we considered $(\lambda_1, \lambda_2, \lambda_3) = (10, 10, 1)$, which represents a large but identical delay times for both treatment groups. Table 3.2 lists the different experimental scenarios of our simulation study. Here, treatment B was considered as the control group and we assume that treatment A always performs better than treatment B since a smaller response is desirable. Scenarios Ia-Ig correspond to a situation when there is a weak correlation between the surrogate and primary endpoints; scenarios IIa-IIe correspond to

a moderate correlation; and scenarios IIIa-IIIe correspond to a strong correlation between the surrogate and primary endpoints.

Normal-Inverse-Wishart prior distribution was placed on (μ_A, Σ_A) and (μ_B, Σ_B), respectively. Conditional on the variance-covariance matrix, the mean vector has a bivariate normal distribution, such that: $\mu_A|\Sigma_A \sim N_2(\mu_0, \Sigma_A/\kappa_0)$, and $\mu_B|\Sigma_B \sim N_2(\mu_0, \Sigma_B/\kappa_0)$. We take the prior expectations of μ_A and μ_B to be both a vector of 0's, but with very large standard deviations. We also consider a weak prior distribution on the variance-covariance matrices Σ_A and Σ_B. Specifically, the following prior parameters are used:

$$\mu_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \qquad R = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \kappa_0 = 0.001, \qquad \upsilon = 4$$

For standard response adaptive randomization procedure, the maximum likelihood estimators were used to calculate the allocation ratio. For the double adaptive biased coin design (DBCD) (3.2), $\alpha = 2$ was used, as suggested by Rosenberger and Hu[Rosenberger and Lachin, 2004]. They found that such a choice of α provides a reasonable trade-off between the randomness and optimality. $5,000$ simulations per scenario was used to evaluate the performance of the proposed response adaptive randomization. To assess the Markov Chain convergence, 3 chains with dispersed initial values were used in each simulation to test whether they all converge to the same target distribution.

3.3.4 Results

To investigate the performance of our proposed response adaptive randomization (proposed RAR) algorithm, the power, allocation proportions to the better treatment group (treatment A), number of subjects assigned to better treatment group (treatment A), and total expected response from all subjects under the proposed allocation algorithm are compared with these under the equal allocation, and the standard response adaptive randomization.

Table 3.7 shows a comparison of equal allocation randomization, standard response adaptive randomization (standard RAR), and proposed response adaptive randomization (proposed RAR) in terms of power for the two-sided two-sample t-test. Both the standard RAR and proposed RAR yield a larger power compared to equal allocation randomization, which is consistent with the conclusion made by Rosenberger and Hu that the doubly-adaptive biased coin design was as powerful or slightly more powerful than the equal allocation procedure [Rosenberger and Hu, 2004]. The power under the proposed RAR algorithm is similar to that under the standard RAR, and it is neither influenced by the delay times in the primary endpoint, nor by the correlation between the primary and surrogate endpoints.

Table 3.8 shows the comparison of the three randomization procedures in terms of the number of subjects assigned to treatment A, along with their standard deviations. As we can see from the table, except for scenarios Ib, IIb, and IIIb, both the standard RAR and proposed RAR assign more subjects to the better treatment group than the equal allocation randomization procedure; compared to the standard RAR, the proposed RAR tends to assign slightly more subjects to treatment A (the

proposed algorithm assigns, on average, 1 to 2 more subjects to the better treatment group, compared with the standard RAR procedure). For scenarios Ib, IIb, and IIIb, where $\mu_{T,A} < \mu_{T,B}$, and $\sigma_{t,A} < \sigma_{t,B}$, the target allocation ratio calculated based on $\rho = \frac{\sqrt{\mu_{T,B}}\sigma_{t,A}}{\sqrt{\mu_{T,B}}\sigma_{t,A}+\sqrt{\mu_{T,A}}\sigma_{t,B}}$ is less than 1/2, however, it is inappropriate to allocate more subjects to the inferior treatment group (treatment B), thus the number of subjects assigned to the better treatment group under both the standard RAR and proposed RAR are the same as equal allocation randomization. Since the equal allocation algorithm does not use the responses during the randomization procedure, its performance is not affected as the delay in the primary endpoint increases. Even the benefit of the standard RAR in terms of assigning more subjects to the better treatment group does not change as the delay increases, the standard deviations of number of subjects assigned to treatment A are getting larger when there is a larger delay in the primary endpoint. The proposed RAR algorithm maintains the benefit of assigning more subjects to the better treatment group, and is not sensitive when the delay times increase. When there is a higher correlation between the surrogate and primary endpoint, the proposed algorithm has a similar or slightly smaller standard deviations of the number of subjects assigned to treatment A.

Table 3.9 shows the observed allocation proportion with standard deviations for the three randomization procedures. As expected, except for scenarios Ib, IIb, and IIIb, both the standard RAR and proposed RAR allocate more than 50% subjects to the better treatment group A. For different delay times in the primary endpoint, the observed allocation proportions under the proposed RAR are similar. The variability in the allocation proportions under the proposed RAR algorithm changes when we

change the correlation between the primary and surrogate endpoints. Specifically, the standard deviation is getting slightly larger when we decrease the correlation between the surrogate and primary endpoints.

Table 3.10 shows the total responses with their standard deviations for the three randomization procedures. Our proposed algorithm results in a comparable or slightly reduction of the total observed responses.

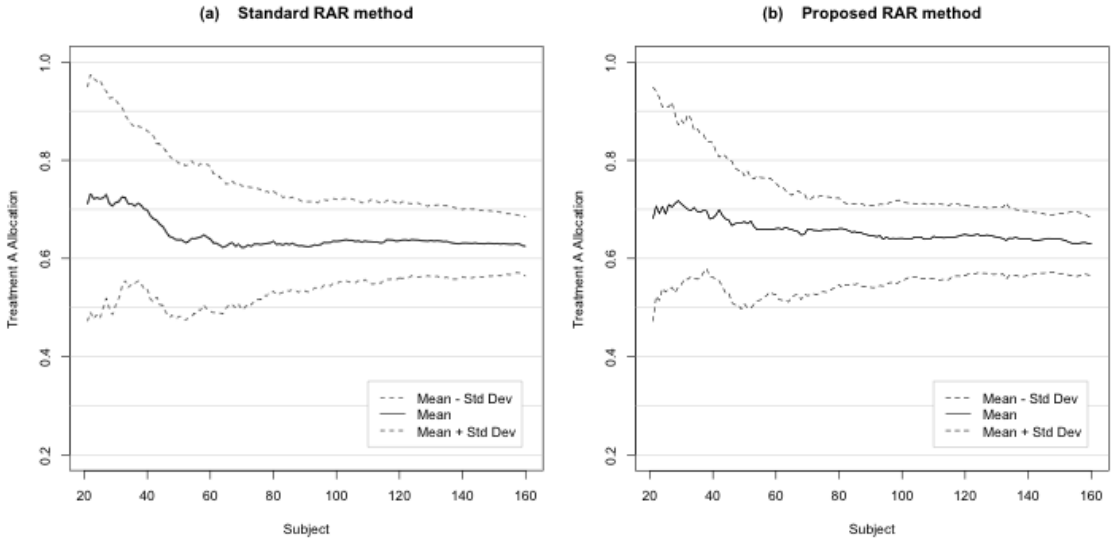
Figure 3.1 and 3.2 compare the performance of the standard RAR and the proposed RAR algorithm with regard to the treatment allocation under different delay parameters. The mean and standard deviation are presented in those figures. In figure 3.1 when the delay parameters are $(\lambda_1, \lambda_2, \lambda_3) = (10, 10, 1)$, the propose RAR algorithm seems to have a smaller variability and stabilize a little bit quicker than the standard RAR, even through the difference is not that huge. In Figure 3.2 when there is a relatively large delay in the primary endpoint where $(\lambda_1, \lambda_2, \lambda_3) = (80, 80, 1)$, we can see that our proposed algorithm reduces the variability and stabilizes much quicker than the standard RAR.

3.4 Discussion

The standard response adaptive randomization procedures, which solely depend on the primary endpoint, are affected by the delays in obtaining the primary outcome measures. When there is a relatively long lag time to observe the primary endpoint, the benefit of standard RAR disappears since few outcome can be used to skew the allocation probability. The proposed algorithm accounts for the informa-

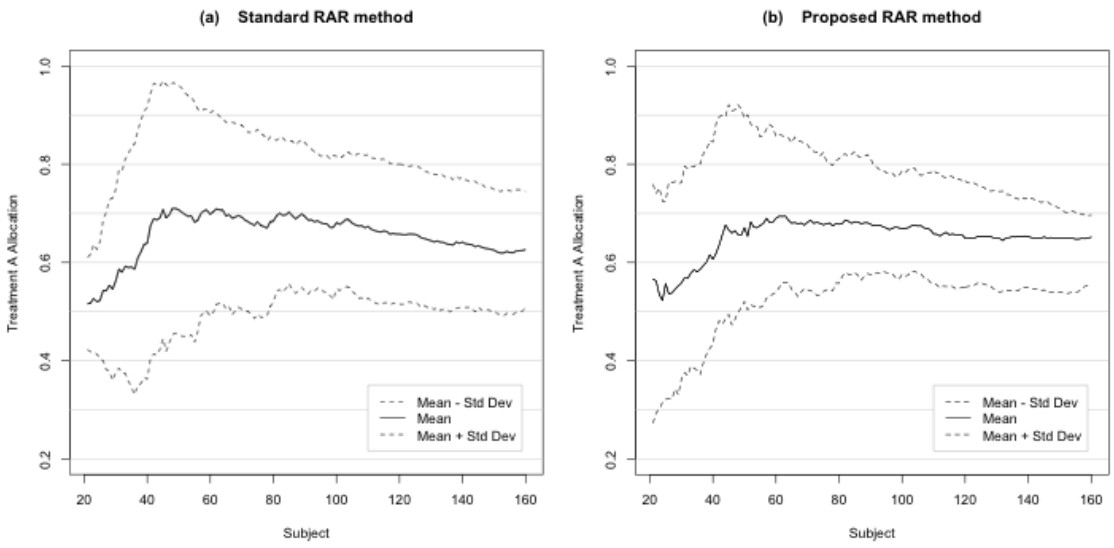
tion from the surrogate endpoint in the randomization procedure. Instead of only using the primary endpoint, the proposed RAR algorithm simultaneously accounts for the surrogate endpoint. Thus, more information will be used to skew the allocation proportion to assign more subjects to the better treatment group under the proposed method. The proposed algorithm results in more subjects in the superior treatment group, while comparable to the standard response adaptive randomization procedures. Under the proposed algorithm, the strength of correlation between the primary and surrogate endpoints does not influence the power and the allocation proportion, but do affect the variability of allocation proportion to the better treatment group.

However, there are some limitations of the proposed algorithm. First, this approach assumes that the surrogate endpoint can be observed immediately after the treatment, which is not always the case. Second, the proposed method we discussed so far only consider one surrogate endpoint. When there are multiple surrogates available, it would be better if we can use those information in the randomization procedure. If all the surrogate endpoints are normally distributed, then we can easily extend the proposed response adaptive randomization from one surrogate endpoint to a multiple surrogate version. Third, the proposed algorithm is only suitable when both the surrogate and primary endpoints are normally distributed. We will consider other distributions of surrogate and primary endpoints in the later chapters.



$N = 160$. Simulations $= 100$.
$\rho_{st} = 0.85, \mu_{T,A} = 13.5, \sigma_{t,A} = 2.5, \mu_{T,B} = 15, \sigma_{t,B} = 4, \mu_{S,A} = 20, \sigma_{s,A} = 4, \mu_{S,B} = 25, \sigma_{s,B} = 3$.
The first 20 subjects were randomized using equal allocation randomization.

Figure 3.1: Comparison of the variation of the standard RAR method and proposed RAR algorithm for handling delayed primary outcomes with $(\lambda_1, \lambda_2, \lambda_3) = (10, 10, 1)$.



$N = 160$. Simulations $= 100$.
$\rho_{st} = 0.85, \mu_{T,A} = 13.5, \sigma_{t,A} = 2.5, \mu_{T,B} = 15, \sigma_{t,B} = 4, \mu_{S,A} = 20, \sigma_{s,A} = 4, \mu_{S,B} = 25, \sigma_{s,B} = 3.$
The first 20 subjects were randomized using equal allocation randomization.

Figure 3.2: Comparison of the variation of the standard RAR method and proposed RAR algorithm for handling delayed primary outcomes with $(\lambda_1, \lambda_2, \lambda_3) = (80, 80, 1)$.

Table 3.2: Parameter setup for different experimental scenario

Scenario	N	$(\lambda_1, \lambda_2, \lambda_3)$	ρ_{st}	Treatment A		Treatment B	
				$(\mu_{T,A}, \sigma_{t,A})$	$(\mu_{S,A}, \sigma_{s,A})$	$(\mu_{T,B}, \sigma_{t,B})$	$(\mu_{S,B}, \sigma_{s,B})$
Ia	90	(10, 10, 1)	0.35	(13, 4)	(20, 4)	(15, 2.5)	(24, 3)
Ib	90	(10, 10, 1)	0.35	(13, 2.5)	(20, 4)	(15, 4)	(24, 3)
Ic	90	(1, 1, 1)	0.35	(13, 4)	(20, 4)	(15, 2.5)	(24, 3)
Id	160	(10, 10, 1)	0.35	(13.5, 4)	(20, 4)	(15, 2.5)	(24, 3)
Ie	178	(10, 10, 1)	0.35	(12.5, 5.8)	(20, 4)	(15, 6)	(24, 3)
If	350	(10, 10, 1)	0.35	(14, 4)	(20, 4)	(15, 2.5)	(24, 3)
IIa	90	(10, 10, 1)	0.6	(13, 4)	(20, 4)	(15, 2.5)	(24, 3)
IIb	90	(10, 10, 1)	0.6	(13, 2.5)	(20, 4)	(15, 4)	(24, 3)
IIc	90	(1, 1, 1)	0.6	(13, 4)	(20, 4)	(15, 2.5)	(24, 3)
IId	160	(10, 10, 1)	0.6	(13.5, 4)	(20, 4)	(15, 2.5)	(24, 3)
IIe	178	(10, 10, 1)	0.6	(12.5, 5.8)	(20, 4)	(15, 6)	(24, 3)
IIIa	90	(10, 10, 1)	0.85	(13, 4)	(20, 4)	(15, 2.5)	(24, 3)
IIIb	90	(10, 10, 1)	0.85	(13, 2.5)	(20, 4)	(15, 4)	(24, 3)
IIIc	90	(1, 1, 1)	0.85	(13, 4)	(20, 4)	(15, 2.5)	(24, 3)
IIId	160	(10, 10, 1)	0.85	(13.5, 4)	(20, 4)	(15, 2.5)	(24, 3)
IIIe	178	(10, 10, 1)	0.85	(12.5, 5.8)	(20, 4)	(15, 6)	(24, 3)
IIIf	350	(10, 10, 1)	0.85	(14, 4)	(20, 4)	(15, 2.5)	(24, 3)

Table 3.3: Summary of power of the three randomization procedures using the two-sided two-sample t-test

Scenario	Sample size (N)	Equal Allocation	Standard RAR	Proposed RAR
Ia	90	0.797	0.829	0.832
Ib	90	0.810	0.807	0.817
Ic	90	0.797	0.830	0.832
Id	160	0.802	0.825	0.836
Ie	178	0.801	0.812	0.801
If	350	0.801	0.816	0.821
IIa	90	0.795	0.835	0.828
IIb	90	0.806	0.806	0.817
IIc	90	0.795	0.837	0.832
IId	160	0.804	0.828	0.830
IIe	178	0.801	0.816	0.806
IIIa	90	0.798	0.839	0.830
IIIb	90	0.808	0.807	0.814
IIIc	90	0.798	0.839	0.828
IIId	160	0.802	0.828	0.826
IIIe	178	0.801	0.815	0.805
IIIf	350	0.798	0.821	0.816

Table 3.4: Summary of average number of subjects assigned to treatment A, with standard deviations (SD) for the three randomization procedures

Scenario	Sample size (N)	Equal Allocation	Standard RAR	Proposed RAR
Ia	90	45 (4.80)	57 (5.06)	58 (5.46)
Ib	90	45 (4.79)	45 (2.64)	45 (2.51)
Ic	90	45 (4.80)	57 (4.65)	58 (5.12)
Id	160	80 (6.34)	100 (6.09)	102 (6.54)
Ie	178	89 (6.64)	93 (5.30)	93 (5.48)
If	350	175 (9.32)	218 (9.18)	219 (9.02)
IIa	90	45 (4.79)	57 (5.07)	58 (5.32)
IIb	90	45 (4.79)	45 (2.58)	45 (2.41)
IIc	90	45 (4.79)	57 (4.63)	58 (5.09)
IIc	160	80 (6.34)	100 (6.17)	101 (6.51)
IId	178	89 (6.64)	93 (5.29)	93 (5.37)
IIIa	90	45 (4.79)	57 (4.99)	58 (5.01)
IIIb	90	45 (4.80)	45 (2.62)	45 (2.30)
IIIc	90	45 (4.79)	57 (4.62)	58 (4.81)
IIId	160	80 (6.24)	100 (6.23)	101 (6.39)
IIIe	178	89 (6.64)	93 (5.28)	93 (5.45)
IIIf	350	175 (9.32)	218 (9.06)	219 (9.08)

Table 3.5: Summary of observed allocation proportion of treatment A, with standard deviations (SD) for the three randomization procedures

Scenario	Sample size (N)	Equal Allocation	Standard RAR	Proposed RAR
Ia	90	0.500 (0.0532)	0.634 (0.0562)	0.646 (0.0607)
Ib	90	0.500 (0.0532)	0.500 (0.0293)	0.501 (0.0279)
Ic	90	0.500 (0.0532)	0.633 (0.0517)	0.649 (0.0569)
Id	160	0.500 (0.0396)	0.627 (0.0381)	0.635 (0.0409)
Ie	178	0.500 (0.0373)	0.522 (0.0297)	0.525 (0.0308)
If	350	0.500 (0.0266)	0.622 (0.0262)	0.625 (0.0258)
IIa	90	0.500 (0.0533)	0.633 (0.0563)	0.645 (0.0592)
IIb	90	0.500 (0.0533)	0.500 (0.0287)	0.501 (0.0267)
IIc	90	0.500 (0.0533)	0.633 (0.0514)	0.647 (0.0565)
IIc	160	0.500 (0.0396)	0.627 (0.0386)	0.634 (0.0407)
IId	178	0.500 (0.0373)	0.522 (0.0297)	0.523 (0.0301)
IIIa	90	0.500 (0.0533)	0.633 (0.0555)	0.641 (0.0557)
IIIb	90	0.500 (0.0533)	0.500 (0.0291)	0.501 (0.0255)
IIIc	90	0.500 (0.0533)	0.633 (0.0513)	0.643 (0.0535)
IIId	160	0.500 (0.0396)	0.626 (0.0390)	0.633 (0.0399)
IIIe	178	0.500 (0.0373)	0.522 (0.0297)	0.524 (0.0306)
IIIf	350	0.500 (0.0266)	0.622 (0.0259)	0.624 (0.0260)

Table 3.6: Summary of observed total responses, with standard deviations (SD) for the three randomization procedures

Scenario	Sample size (N)	Equal Allocation	Standard RAR	Proposed RAR
Ia	90	1260 (33.47)	1236 (35.95)	1233 (36.30)
Ib	90	1260 (33.18)	1260 (32.16)	1261 (31.59)
Ic	90	1260 (33.47)	1235 (35.97)	1234 (35.46)
Id	160	2280 (44.04)	2250 (46.95)	2248 (46.30)
Ie	178	2448 (81.21)	2438 (79.61)	2437 (79.86)
If	350	5075 (63.38)	5032 (67.39)	5031 (67.40)
IIa	90	1260 (33.39)	1236 (36.15)	1234 (36.23)
IIb	90	1260 (33.17)	1260 (32.26)	1261 (31.75)
IIc	90	1260 (33.39)	1236 (35.78)	1234 (35.64)
IId	160	2280 (44.26)	2249 (46.77)	2248 (46.44)
IIe	178	2448 (81.20)	2438 (79.56)	2437 (79.66)
IIIa	90	1260 (33.28)	1235 (36.22)	1234 (36.10)
IIIb	90	1260 (33.10)	1260 (32.37)	1261 (31.80)
IIIc	90	1260 (33.28)	1235 (35.93)	1234 (35.36)
IIId	160	2280 (44.04)	2249 (46.82)	2249 (46.13)
IIIe	178	2448 (81.21)	2437 (79.79)	2437 (79.80)
IIIf	350	5075 (63.05)	5032 (66.39)	5032 (67.09)

Table 3.7: Summary of power of the three randomization procedures using the two-sided two-sample t-test

Scenario	Sample size (N)	Equal Allocation	Standard RAR	Proposed RAR
Id	160	0.802	0.833	0.834
Ie	178	0.801	0.812	0.803
If	350	0.801	0.816	0.816
IId	160	0.804	0.836	0.830
IIe	178	0.801	0.816	0.801
IIId	160	0.802	0.828	0.825
IIIe	178	0.801	0.815	0.804
IIIf	350	0.798	0.821	0.816

Table 3.8: Summary of average number of subjects assigned to treatment A, with standard deviations (SD) for the three randomization procedures

Scenario	Sample size (N)	Equal Allocation	Standard RAR	Proposed RAR
Id	160	80 (6.34)	100 (6.04)	102 (6.48)
Ie	178	89 (6.64)	93 (5.30)	93 (5.48)
If	350	175 (9.32)	218 (9.18)	219 (9.10)
IIc	160	80 (6.34)	100 (6.09)	101 (6.46)
IId	178	89 (6.64)	93 (5.29)	93 (5.35)
IIId	160	80 (6.24)	100 (6.23)	101 (6.32)
IIIe	178	89 (6.64)	93 (5.28)	93 (5.46)
IIIf	350	175 (9.32)	218 (9.06)	219 (9.08)

Table 3.9: Summary of observed allocation proportion of treatment A, with standard deviations (SD) for the three randomization procedures

Scenario	Sample size (N)	Equal Allocation	Standard RAR	Proposed RAR
Id	160	0.500 (0.0396)	0.627 (0.0378)	0.634 (0.0405)
Ie	178	0.500 (0.0373)	0.522 (0.0297)	0.524 (0.0308)
If	350	0.500 (0.0266)	0.622 (0.0262)	0.624 (0.0260)
IIc	160	0.500 (0.0396)	0.627 (0.0381)	0.634 (0.0404)
IId	178	0.500 (0.0373)	0.522 (0.0297)	0.523 (0.0301)
IIId	160	0.500 (0.0396)	0.626 (0.0390)	0.633 (0.0394)
IIIe	178	0.500 (0.0373)	0.522 (0.0297)	0.524 (0.0307)
IIIf	350	0.500 (0.0266)	0.622 (0.0259)	0.624 (0.0260)

Table 3.10: Summary of observed total responses, with standard deviations (SD) for the three randomization procedures

Scenario	Sample size (N)	Equal Allocation	Standard RAR	Proposed RAR
Id	160	2280 (44.04)	2249 (46.38)	2248 (46.50)
Ie	178	2448 (81.21)	2438 (79.61)	2437 (80.15)
If	350	5075 (63.38)	5032 (67.39)	5032 (67.37)
IId	160	2280 (44.26)	2249 (46.32)	2248 (46.51)
IIe	178	2448 (81.20)	2438 (79.56)	2438 (79.83)
IIId	160	2280 (44.04)	2249 (46.82)	2248 (46.17)
IIIe	178	2448 (81.21)	2437 (79.79)	2437 (79.99)
IIIf	350	5075 (63.05)	5032 (66.39)	5032 (67.09)

Chapter 4

Response Adaptive Design for Clinical Trials with Time-to-Event Outcomes using Surrogate Endpoint

When the outcome of interest is time to event, the procedure used in the previous chapter does not apply exactly. In this chapter, we are going to propose a response adaptive randomization for clinical trials with survival outcomes, which will simultaneously account for the surrogate endpoint. Almost always, the survival outcomes are delayed. As a result, we are not able to adapt the design. A surrogate, which can be observed early, is expected to help in making the design adapt toward the right direction. The exponential model is the most fundamental parametric model

and is commonly used in survival analysis. Moreover, the exponential distribution usually leads to closed form theoretical results [Zhang and Rosenberger, 2007]. Thus, we will assume the primary time-to-event outcome has an exponential distribution throughout this chapter.

4.1 RAR using primary outcome only

Consider a simple clinical trial with two treatment groups, Treatment A and Treatment B. Let n_A and n_B be the number of patients in each treatment group, and $n_A + n_B = n$. Suppose that the the primary endpoint of interest is time to event outcome. Specifically, suppose T_{ik} is the survival time for the ith patient in group k, and follows an exponential distribution with mean θ_k, $k = A, B$. Furthermore, assume that the survival times are subject to right censoring. Let C_i be the censoring time for the ith patient, and is assumed to be independent of the survival time T_{ik}. Then for the ith patient in treatment k, $t_{ik} = \min(T_{ik}, C_i)$ is the observed or censored survival time with corresponding indicator variable δ_{ik}, where $\delta_{ik} = 1$ if the ith patient in treatment k is observed, and $\delta_{ik} = 0$ if that patient is censored.

As we mentioned in section 2.2, the general technique to obtain an optimal target allocation is to solve the following optimization problem:

$$\begin{cases} \min\limits_{n_A/n_B} & n_A\Psi_A + n_B\Psi_B \\ \text{s.t.} & \frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B} \leq \kappa \end{cases} \tag{4.1}$$

where μ_k and σ_k^2, $k = A, B$, are the mean and finite variance of the primary response,

and Ψ_k is a positive function of (μ_k, σ_k), which will be different for different goal of the trial, and κ is a constant. Then the optimal allocation ratio would be

$$\rho = \frac{\sigma_A\sqrt{\Psi_B}}{\sigma_A\sqrt{\Psi_B} + \sigma_B\sqrt{\Psi_A}} \tag{4.2}$$

Assume that $\epsilon_k = E(\delta_{ik})$ is the non-censoring proportion for patients in each treatment group, and is fixed for patients in the same treatment group. Then under the above setup, Zhang and Rosenberger (2007) proposed an optimal allocation proportion by minimizing the total expected hazard [Zhang and Rosenberger, 2007]. The optimal allocation proportion can be obtained from equation (4.1) by taking $\Psi_k = \theta_k^{-1}$ and $\sigma_k^2 = \frac{\theta_k^2}{\epsilon_k}$:

$$\begin{cases} \min\limits_{n_A/n_B} & n_A\theta_A^{-1} + n_B\theta_B^{-1} \\ \text{s.t.} & \frac{\theta_A^2}{n_A\epsilon_A} + \frac{\theta_B^2}{n_B\epsilon_B} \leq \kappa \end{cases} \tag{4.3}$$

Solving the above optimization problem, we can get the optimal allocation proportion of patients who are assigned to treatment A as

$$\rho = \frac{\sqrt{\theta_A^3\epsilon_B}}{\sqrt{\theta_A^3\epsilon_B} + \sqrt{\theta_B^3\epsilon_A}} \tag{4.4}$$

The non-censoring proportion ϵ_k will depend on the censoring scheme used in the trial. In this chapter, we will introduce the censoring scheme proposed by Rosenberger and Seshaiyer (1997) and assume their censoring scheme throughout this chapter [Rosenberger and Seshaiyer, 1997]. Suppose the trial has a duration $D > 0$, and a recruitment period of length $R > 0$ and $R < D$. Patients enter the trial

sequentially and patient arrival times are independent and follow an uniform distribution on $[0, R]$. Meanwhile, the censoring time C is independent from the survival time and is assumed to be uniformly distributed over $[0, D]$. Patients who do not respond by the end of the trial are considered as administratively censored. Then for the ith patient, we can observe $t_{ik} = \min(T_{ik}, C_i, D - R)$ and $\delta_{ik} = 1$ if $t_{ik} = T_{ik}$ and $\delta_{ik} = 0$ otherwise. Zhang and Rosenberger (2007) found that the non-censoring proportion $\epsilon_k = \Pr(\delta_{ik} = 1)$ under the above censoring scheme has the following form [Zhang and Rosenberger, 2007]

$$\epsilon_k = 1 - \frac{\theta_k}{D} + \exp\left(-\frac{D}{\theta_k}\right)\frac{\theta_k}{DR}\left\{\exp\left(\frac{R}{\theta_k}\right)(2\theta_k - R) - 2\theta_k\right\} \tag{4.5}$$

Under equation (4.5), Sverdlov et al. (2011) have shown that for fixed values of D and R, the non-censoring proportion ϵ_k is monotonically decreasing when the mean parameter θ_k increases [Sverdlov et al., 2011]. This is easy to understand that the longer the expected survival time, the less likely to observe an event (for example death) before censoring, thus a smaller non-censoring proportion will be expected.

Then for the standard response adaptive randomization (RAR) where we only consider the primary endpoint, the maximum likelihood estimator of the θ_k can be obtained from the data and thus the optimal allocation proportion ρ can be estimated every time before assigning the next patient.

4.2 Proposed method

4.2.1 The finite mixture framework

A finite mixture model is a convex combination of two or more probability density functions. It enriches the set of probability models by adding finite mixtures (or weighted sums) of other standard distributions [Deb et al., 2008]. In general, the density function of a m-component finite mixture is:

$$f(x) = \sum_{j=1}^{m} p_j f_j(x) \tag{4.6}$$

where $0 < p_j < 1$ is weighting factor, $\sum_{j=1}^{m} p_j = 1$, and $f_j(x)$ is the $p.d.f.$ of the jth component (e.g. Gaussian, Exponential, Weibull, etc).

In this chapter, we are going to consider a finite mixture of m exponential distributions, which can also be called as a hyper-exponential distribution. Let $X_1, \ldots, X_n$ be a random sample from the hyper-exponential distribution, then the marginal distribution of X can be expressed as:

$$f(x) = \sum_{j=1}^{m} p_j \frac{1}{\theta_j} e^{-\frac{x}{\theta_j}} \tag{4.7}$$

The hyper-exponential cumulative distribution function for X, derived from equation (4.7), is

$$F(x) = \sum_{j=1}^{m} p_j (1 - e^{-\frac{x}{\theta_j}}) \tag{4.8}$$

And the survival function for X, derived from (4.8), is

$$S(x) = 1 - F(x) = \sum_{j=1}^{m} p_j e^{-\frac{x}{\theta_j}} \tag{4.9}$$

Then the expectation of X derived from hyper-exponential distribution in (4.7), can be expressed as below:

$$E(X) = \sum_{j=1}^{m} p_j \theta_j \tag{4.10}$$

4.2.2 Specifying the design parameters

Let n_k be the number of subjects in treatment k, where k is the treatment indicator and $k = A$ or B. Suppose that there is a surrogate endpoint that has been validated and this surrogate endpoint has m categories. Specifically, if the ith subject in treatment k has a surrogate endpoint in the jth category, $j = 1, 2, \ldots, m$, then denote this by $S_{k,j,i} = 1$ and $S_{k,r,i} = 0$ for $1 \leq r \leq m, r \neq j$. Let $\boldsymbol{S_{k,i}} = (S_{k,1,i}, \ldots, S_{k,m,i})$ be a vector, and we assume that the vectors are independent and identically distributed across $i = 1, \ldots, n_k$, and each follows a multinomial distribution. Furthermore, suppose the primary endpoint is a survival time that follows an exponential distribution, and we denote $T_{k,i}$ as the survival time for subject i in treatment k. Then conditional on the surrogate endpoint being in the jth category, we assume $T_{k,i}$ follows an exponential distribution with mean parameter $\theta_{k,j}$, $j = 1, \ldots, m$. Then under the above assumptions, the primary endpoint $T_{k,i}$, $i = 1, 2, \ldots, n_k$, has a mixture of

exponential distribution. The model can be presented as below:

$$\begin{aligned} &k = A, B; \\ &(S_{k,1,i}, S_{k,2,i}, \ldots, S_{k,m,i}) \sim \text{Multi}(1, p_{k,1}, p_{k,2}, \ldots, p_{k,m}); \\ &T_{k,i} \sim \sum_{j=1}^{m} p_{k,j} \text{Exp}(\theta_{k,j}) \end{aligned} \tag{4.11}$$

where $p_{k,j}, j = 1, \ldots, m$ is the probability of a subject in treatment k having a surrogate endpoint in the jth category, and $\sum_{j=1}^{m} p_{k,j} = 1$, $m \geq 2$. Based on the model setup in (4.11), the marginal distribution of survival time (primary endpoint) can be expressed as:

$$f(t_{k,i}) = \sum_{j=1}^{m} p_{k,j} \frac{1}{\theta_{k,j}} e^{-\frac{t_{k,i}}{\theta_{k,j}}} \tag{4.12}$$

The survival function for the primary endpoint given the surrogate endpoint can be written as

$$S(t_{k,i}) = \sum_{j=1}^{m} p_{k,j} e^{-\frac{t_{k,i}}{\theta_{k,j}}} \tag{4.13}$$

The expectation of survival time for each group can be written as:

$$\theta_k = \sum_{j=1}^{m} p_{k,j} \theta_{k,j} \tag{4.14}$$

In this chapter, we are only going to consider the surrogate endpoint with two categories, that is $m = 2$. So the mixture model will only have 2 components. Suppose that associated with $T_{k,i}$, there is a censoring time C_i, and C_i is assumed to be independent from $T_{k,i}$. Thus all subjects may have an event or be censored. Then

for each subject, we will observe a pair of $(t_{k,i}, \delta_{k,i})$, where $t_{k,i} = \min(T_{k,i}, C_i)$ is the observed time and $\delta_{k,i}$ is an indicator if the event with $\delta_{k,i} = 1$ if $t_{k,i} = T_{k,i}$, and $\delta_i = 0$ if the ith subject in treatment k is censored. If we let $\phi = (\theta_{k,1}, \theta_{k,2}, p_{k,1}, p_{k,2})$ be a set of unknown parameters, then the likelihood function for the observed data can be written as

$$L(\phi|\text{Data}) = \prod_{k=A}^{B} \prod_{i=1}^{n_k} \{f(t_{k,i})\}^{\delta_{k,i}} \{S(t_{k,i})\}^{1-\delta_{k,i}} \tag{4.15}$$

where $f(t_i)$ is the distribution of survival time for subject i as defined in equation (4.12), and $S(t_i)$ is the corresponding survival function as equation (4.13).

4.3 Model fitting using MCMC

To make inference for the unknown parameters $\theta_{k,j}$ and p_j, a Bayesian approach will be introduced. Due to the complexity of the model, a Bayesian approach is deemed to be appropriate. Posterior inference for mixture models can be performed via the Markov Chain Monte Carlo simulation. Casella et al. (2002) mentioned that before the Markov Chain Monte Carlo (MCMC) was introduced, there was no satisfactory way to compute the Bayes estimators for mixture model [Casella et al., 2002, Marin et al., 2005]. We are going to use the Gibbs sampler of Gelfand and Smith (1990) to get the Bayes estimators of the unknown parameters.

4.3.1 Label switching

Label-switching is a common issue in Bayesian estimation of mixture models. Consider the mixture model in (4.12), where the mixture model has a finite number of components (m). If the prior distributions on $(\theta_1, \ldots, \theta_m)$ are exchangeable, then all the marginals on the θ_j's, $j = 1, \ldots, m$ are identical, thus posterior distribution of θ_j's are invariant under permutations of the indices of the components. This means that we cannot distinguish θ_1 from θ_2, θ_1 from θ_3, and so on, from the likelihood, since they are exchangeable. This identifiability feature is crucial for both Bayesian inference and computational issues [Marin et al., 2005].

There have been many suggestions as to how to deal with the label switching problem. One solution is to use artificial identifiability constraints on the parameters to break the symmetry in the likelihood [McLachlan and Peel, 2004]. For example, if we go back to the mixture model in (4.12), a possible constraint is to order the mean parameters, such that $\theta_1 < \theta_2 < \cdots < \theta_m$. This approach performs well when the number of components is small. However, for a large number of components, Celeux et al. pointed out that identifiability constraints have a consequence on the posterior distributions, that it may lead to very poor estimates of the distribution in the end. In that case, some other approach can be used to handle the label switching issue, such as relabelling algorithms. We are not going to talk in detail of relabelling algorithms in this chapter.

4.3.2 Prior distribution

The Dirichlet distribution is a multivariate generalization of the beta distribution. Similar to the beta distribution which can be used to measure the uncertainty about two positive numbers that must sum to 1, the Dirichlet distribution can be used to measure the uncertainty about m positive numbers that must sum to 1 [Escoto, 2013]. The Dirichlet distributions are commonly used as prior distributions in Bayesian statistics. Moreover, just as the beta distribution is a conjugate prior of the Bernoulli distribution, the Dirichlet distribution is the conjugate prior of the categorical/ multinomial distribution.

In this chapter, we will consider the Dirichlet distribution as a prior for the surrogate endpoint. Now go back to the mixture model in section (4.2.2), where the surrogate endpoint follows a multinomial distribution, $(S_{k,1,i}, S_{k,2,i}, \ldots, S_{k,m,i}) \sim \text{Multi}(1, p_{k,1}, p_{k,2}, \ldots, p_{k,m})$ for $k = A, B$ and $i = 1, \ldots, n_k$, where $p_{k,j}$ is the probability of a subject in treatment k has a surrogate endpoint in the jth category. Thus in the Bayesian framework, we can assume that the vector of probability parameters $(p_{k,1}, \ldots, p_{k,m})$ has a Dirichlet prior with parameters $(\gamma_{k,1}, \ldots, \gamma_{k,m})$, such that $(p_{k,1}, \ldots, p_{k,m}) \sim \text{Dir}(\gamma_{k,1}, \ldots, \gamma_{k,m})$. And the density function will be given by

$$\begin{aligned} \pi(p_{k,1}, \ldots, p_{k,m} | \gamma_{k,1}, \ldots, \gamma_{k,m}) &= \frac{1}{\text{B}(\gamma_{k,1}, \ldots, \gamma_{k,m})} \prod_{j=1}^{m} p_{k,j}^{\gamma_{k,j}-1} \\ &\propto \prod_{j=1}^{m} p_{k,j}^{\gamma_{k,j}-1} \end{aligned} \tag{4.16}$$

where $\text{B}(\gamma_{k,1}, \ldots, \gamma_{k,m}) = \frac{\prod_{j=1}^{m} \Gamma(\gamma_{k,j})}{\Gamma(\sum_{j=1}^{m} \gamma_{k,j})}$. If we let $\gamma_{k,0} = \sum_{j=1}^{m} \gamma_{k,j}$, then the expectation

of each probability parameter is equal to $\mathrm{E}[p_{k,j}] = \frac{\gamma_{k,j}}{\gamma_{k,0}}$. Moreover, $\gamma_{k,0}$ reflects the uncertainty of the mixed exponential prior distribution. Specifically, the larger the sum of the parameters $\gamma_{k,0}$, the more certain we are of the true weights (or probability parameters) [Escoto, 2013].

The Gamma distribution is a conjugate prior to the exponential distributions. However, as we mentioned in section (4.3.1), if we put an inverse gamma prior on each of mean parameters $(\theta_{k,1}, \ldots, \theta_{k,m})$, then their marginal distributions are identical, and their posterior distributions are invariant. To solve the label switching problem, a semi-conjugate prior will be chosen for the mean parameters $(\theta_{k,1}, \ldots, \theta_{k,m})$, that is put an identifiability constraint on $\theta_{k,1}, \ldots, \theta_{k,m}$, such that $\theta_{k,1} > \theta_{k,2} > \ldots > \theta_{k,m}$. Remember that we will only consider that the surrogate endpoint has 2 categories $(m = 2)$. Then the constraint will be $\theta_{k,2} < \theta_{k,1}$. Furthermore, if we assume both $\theta_{k,1}$ and $\theta_{k,2}$ have an Inverse-Gamma prior with parameter $(\alpha_{k,1}, \beta_{k,1})$ and $(\alpha_{k,2}, \beta_{k,2})$, respectively, then the semi-conjugate prior for $\theta_{k,1}$ and $\theta_{k,2}$ can be written as

$$\begin{aligned}\pi(\theta_{k,1}, \theta_{k,2} | \alpha_{k,1}, \beta_{k,1}, \alpha_{k,2}, \beta_{k,2}) &= \frac{\beta_{k,1}}{\Gamma(\alpha_{k,1})} \theta_{k,1}^{-\alpha_{k,1}-1} \exp(-\frac{\beta_{k,1}}{\theta_{k,1}}) \\ &\times \frac{\beta_{k,2}}{\Gamma(\alpha_{k,2})} \theta_{k,2}^{-\alpha_{k,2}-1} \exp(-\frac{\beta_{k,2}}{\theta_{k,2}}) I(\theta_{k,2} < \theta_{k,1}) \\ &\propto \theta_{k,1}^{-\alpha_{k,1}-1} \theta_{k,2}^{-\alpha_{k,2}-1} \exp(-\frac{\beta_{k,1}}{\theta_{k,1}} - \frac{\beta_{k,2}}{\theta_{k,2}}) I(\theta_{k,2} < \theta_{k,1}).\end{aligned} \tag{4.17}$$

Combining the likelihood function in (4.15) and the above prior distributions for $(p_{k,1}, p_{k,2})$ in (4.16) and $(\theta_{k,1}, \theta_{k,2})$ in (4.17), we will get the complete Bayesian model and thus the posterior distribution of the set of unknown parameters $\phi =$

$(\theta_{k,1}, \theta_{k,2}, p_{k,1}, p_{k,2})$ given the data, which can be expressed as below:

$$\pi(\theta_{k,1}, \theta_{k,2}, p_{k,1}, p_{k,2}) \propto L(\phi|\text{Data})\pi(p_{k,1}, p_{k,2})\pi(\theta_{k,1}, \theta_{k,2}) \tag{4.18}$$

4.3.3 Estimation procedure

To calculate the allocation rate, we need to estimate the marginal mean survival time for each treatment group. As we mentioned before, the unknown parameters will be estimated through a Bayesian model. Specifically, we will use the Markov Chain Monte Carlo (MCMC) methods through JAGS and R to fit our proposed model and then obtain samples from the posterior distribution. JAGS (Just Another Gibbs Sampler) is a well established statistical program for analysis of Bayesian hierarchical models using Markov Chain Monte Carlo (MCMC) simulation [Plummer et al., 2003].

In section (4.2.2), we assume that the surrogate endpoint has a multinomial distribution, and conditional on the surrogate endpoint being in the jth category, the primary outcome $T_{k,i}$ follows an exponential distribution with mean parameter $\theta_{k,j}$:

$$\begin{aligned}
&k = A, B;\\
&(S_{k,1,i}, S_{k,2,i}, \ldots, S_{k,m,i}) \sim \text{Multi}(1, p_{k,1}, p_{k,2}, \ldots, p_{k,m});\\
&T_{k,i} \sim \sum_{j=1}^{m} p_{k,j}\text{Exp}(\theta_{k,j})
\end{aligned}$$

As we mentioned in section (4.3.2), a semi-conjugate prior was chosen for the mean parameters $(\theta_{k,1}, \ldots, \theta_{k,m})$ to solve the label switching problem. To make the model

easily implemented in JAGS, we will re-write the prior for the mean parameters $(\theta_{k,1}, \theta_{k,2})$ as below:

$$\begin{aligned} \theta_{k,2} &\sim \text{IG}(\alpha_{k,2}, \beta_{k,2}); \\ \Delta_k &\sim \text{IG}(\alpha_{k,0}, \beta_{k,0}); \\ \theta_{k,1} &= \Delta_k + \theta_{k,2}. \end{aligned} \tag{4.19}$$

This prior setup can make sure that we put an identifiability constraint on the mean parameters, such that $\theta_{k,1} > \theta_{k,2}$. Then the Bayesian model for our mixture of two exponential distributions can be present below:

$$\begin{aligned} & k = A, B; \\ & (S_{k,1,i}, S_{k,2,i}) \sim \text{Multi}(1, p_{k,1}, p_{k,2}), \qquad \text{for} i = 1, 2, \ldots, n_k; \\ & T_{k,i} \sim \sum_{j=1}^{2} p_{k,j} \text{Exp}(\theta_{k,j}); \\ & (p_{k,1}, p_{k,2}) \sim \text{Dir}(\gamma_{k,1}, \gamma_{k,2}); \\ & \theta_{k,2} \sim \text{IG}(\alpha_{k,2}, \beta_{k,2}); \\ & \Delta_k \sim \text{IG}(\alpha_{k,0}, \beta_{k,0}); \\ & \theta_{k,1} = \Delta_k + \theta_{k,2}. \end{aligned} \tag{4.20}$$

where $\text{Exp}(\cdot)$ is the exponential distribution, $\text{Dir}(\cdot)$ is the Dirichlet distribution, and $\text{IG}(\cdot)$ is the inverse gamma distribution. Then we will use the JAGS to obtain the posterior mean of unknown parameters $p_{k,1}, p_{k,2}, \theta_{k,1}, \theta_{k,2}$, and denote them as $\tilde{p}_{k,1}, \tilde{p}_{k,2}, \tilde{\theta}_{k,1}, \tilde{\theta}_{k,2}$, respectively. Then, we will use equation (4.10) to estimate the

mean survival time for each treatment group:

$$\tilde{\theta}_k = \tilde{p}_{k,1}\tilde{\theta}_{k,1} + \tilde{p}_{k,2}\tilde{\theta}_{k,2}$$

and plug these estimates in equation (4.5) to get the estimate of the non-censoring proportion:

$$\tilde{\epsilon}_k = 1 - \frac{\tilde{\theta}_k}{D} + \exp\left(-\frac{D}{\tilde{\theta}_k}\right)\frac{\tilde{\theta}_k}{DR}\left\{\exp\left(\frac{R}{\tilde{\theta}_k}\right)(2\tilde{\theta}_k - R) - 2\tilde{\theta}_k\right\}$$

4.4 Algorithm of the design

The proposed algorithm for response adaptive randomization for survival primary outcomes is described below:

S1. To begin with the procedure, first an equal randomization with a prefixed number of subjects $2m_0$ will be performed. Equal randomization is a useful way to obtain initial parameter estimates that are required in a sequential estimation procedure, such as the DBCD procedure. However, the prefixed number os subjects in the equal randomization procedure is not clear and most time is arbitrary without any statistical justification [Xu and Yin, 2014]. Xu and Yin (2014) selected the number of subjects in the equal randomization stage large enough to make sure there is a treatment difference before they moved to the adaptive randomization. Nowacki et al. (2015) chose $5\% - 10\%$ of the total sample size in the equal randomization procedure [Nowacki et al., 2015].

S2. Before allocating the ith subject, update the following information based on the first $i-1$ subjects:

$n_{A,i}$: current total number of subjects assigned to treatment A;

$n_{B,i}$: current total number of subjects assigned to treatment B.

S3. Calculate the Bayes estimators of the unknown parameters through the Bayesian model based on the accumulate data:

$\tilde{\theta}_A = \tilde{p}_{1,A}\tilde{\theta}_{1,A} + \tilde{p}_{2,A}\tilde{\theta}_{2,A}$: posterior estimate of marginal mean survival time for A;

$\tilde{\theta}_B = \tilde{p}_{1,B}\tilde{\theta}_{1,B} + \tilde{p}_{2,B}\tilde{\theta}_{2,B}$: posterior estimate of marginal mean survival time for B;

$\tilde{\epsilon}_A$: estimate of non-censoring proportion for subjects in treatment A;

$\tilde{\epsilon}_B$: estimate of non-censoring proportion for subjects in treatment A.

S4. Calculate the current treatment A allocation proportion and the current treatment A target allocation proportion:

$r_{A,i} = \frac{n_{A,i}}{n_{A,i}+n_{B,i}}$: current observed proportion of subjects assigned to treatment A;

$\hat{\rho} = \frac{\sqrt{\tilde{\theta}_A^3\tilde{\epsilon}_B}}{\sqrt{\tilde{\theta}_A^3\tilde{\epsilon}_B}+\sqrt{\tilde{\theta}_B^3\tilde{\epsilon}_A}}$: current estimate of target allocation proportion to treatment A.

S5. Apply both the current observed treatment A allocation proportion and the current treatment A target allocation proportion calculated in the previous step

to the DBCD procedure to calculate the treatment A allocation probability for the ith subject:

$$Prob(TrtA) = \frac{\hat{\rho}(\frac{\hat{\rho}}{r_{A,i}})^{\alpha}}{\hat{\rho}(\frac{\hat{\rho}}{r_{A,i}})^{\alpha} + (1-\hat{\rho})(\frac{1-\hat{\rho}}{1-r_{A,i}})^{\alpha}}$$

where α is a nonnegative number that reflects the desired degree of randomization.

S6. Randomize the next subject using this treatment A allocation probability $Prob(TrtA)$.

S7. Repeat steps 2 - 6 until reaching the predetermined sample size.

4.5 Simulation study

4.5.1 Simulation targets

In this section, we will conduct a number of simulations to evaluate the performance of our proposed response adaptive randomization for clinical trials with time-to-event primary outcome and binary surrogate endpoint under different clinical scenarios. Specifically, our proposed algorithm will be compared with two other randomization procedures: (i) equal randomization procedure for which each subjects is assigned to either treatment A or treatment B with probabilities $(0.5, 0.5)$ and (ii) the standard response adaptive randomization for survival outcome proposed by Zhang and Rosenberger in 2007, where the allocation proportion is updated solely based on the

primary endpoint.

As recommended by Sverdlov et al. (2011), several characteristics will be compared between the proposed algorithm and the other two randomization procedures to evaluate the performance of the new randomization procedure: (1) simulated allocation proportions: larger proportion of subjects are expected to be assigned to more efficacious treatment group; (2) the power for testing equality of treatment effects: the power under the proposed algorithm is expected to be at least as powerful as the other two randomization procedure; (3) average number of patients on the superior treatment group; (4) average number of events (deaths); and (5) total observed survival time [Sverdlov et al., 2011]. So in our simulation study, we are interested in: 1) how different response rate in the surrogate endpoint affect the simulation results, in terms of the five characteristics we talked before; 2) the effect of delay (censoring proportion) in the primary endpoint; and 3) how different sample sizes affect the simulation results.

4.5.2 Sampling method

For each clinical scenario, the sample size was chosen such that the equal randomization procedure has at least 80% power and 0.05 significance level of the two-sided Wald test for testing $H_0 : \theta_A = \theta_B$:

$$Z = \frac{\hat{\theta}_A - \hat{\theta}_B}{\sqrt{\hat{\theta}_A^2/r_A + \hat{\theta}_B^2/r_B}}$$

where $\hat{\theta}_A$ and $\hat{\theta}_B$ are the maximum likelihood estimators of mean survival time for each treatment group, and r_A and r_B are the total number of failures for each treatment group.

The surrogate endpoint was sampled from a two dimensional multinomial distribution (or equally the binomial distribution). And then the primary endpoint was sampled from a mixture of two exponential distributions based on the sampled surrogate endpoint. As we mentioned before, Rosenberger and Seshaiyer's (1997) censoring scheme will be assumed throughout this chapter. So for both treatment A and treatment B, subject arrival time was generated from a uniform distribution over $[0, R]$, and the censoring time was sampled from a uniform distribution over $[0, D]$. For convergence diagnostics, then Gelman-Rubin diagnostic was used as a numerical support for the convergence of the chains. The Gelman-Rubin approach is, for each parameter or quantity of interest, to compute the variance of the simulations from each chain, to average these within-chain variances, and compare this to the variances of all the chains mixed together [Gelman et al., 2011]. Basically, Gelman-Rubin measures whether there is a significant difference between the variance within several chains and the variance between several chains by a value that is called "scale reduction factors".

4.5.3 Simulation settings

As in section 4.1, Rosenberger and Seshaiyer's (1997) censoring scheme with recruitment period $R = 55$ and duration $D = 96$ was considered. This setup was chosen to match the experimental setting of a head and neck cancer trial reported by Fountzi-

las et al. [Fountzilas et al., 2004]. And this setup was also commonly used in a lot of simulation studies of response adaptive randomization for survival outcome. The surrogate endpoint for each treatment group was generated from a multinomial distribution Multi$(1, p_{k,1}, p_{k,2})$. Different combinations of the response rates of surrogate endpoint for each treatment group was considered. The primary survival time was generated from the mixture of two exponential distributions. We will investigate the performance of our proposed algorithm under different expected survival times. Table 4.1 lists the different experimental scenario of our simulation study. Treatment B was considered as the control group and we assume that treatment A always performs better than treatment B. Furthermore, we assume that the mean survival time for subjects having a surrogate endpoint in the first category is greater than the mean survival time if having a surrogate endpoint in the second category. Scenarios Ia-If correspond to a situation when subjects in treatment A have a higher rate of having a surrogate endpoint in the first category than in treatment B, and the mean survival time in treatment A is greater than that in treatment B. IIa-IIe correspond to a situation when subjects in treatment A have the same rate of having a surrogate endpoint in the first category as in treatment B, but treatment A has a higher survival time compared to treatment B. And scenarios IIIa-IIIc correspond to a situation when subjects in treatment A have a higher rate of having a surrogate endpoint in the first category than in treatment B, but the mean survival times for subjects having a surrogate endpoint in the jth category for both treatment groups are the same. We should note that a larger expected survival times corresponds to a situation when there is a larger proportion of censoring, which can also be consid-

ered as a larger proportion of delays. Table 4.1 also lists the proportion of censoring (or delay) for each treatment group under different clinical scenarios. For example, under scenario Ia, the proportion of non-censoring for treatment A is 0.49, while 0.80 for treatment B. When the mean survival time for treatment A decreases, the proportion of non-censoring decreases as well.

The Dirichlet distribution was considered as the prior distribution for the surrogate endpoint. For both treatment A and treatment B, we assume that $(p_{k,1}, p_{k,2}) \sim \text{Dir}(0.5, 0.5)$. So $\gamma_{k,0} = \gamma_{k,1} + \gamma_{k,2} = 1$. Remember in section 4.3.2 we mentioned that $\gamma_{k,0}$ reflects the uncertainty of the mixed exponential prior distribution, and the larger the $\gamma_{k,0}$ is, the more certain we are of the true values of $p_{k,1}$ and $p_{k,2}$. So under this parameterization for the Dirichlet distribution, we assume that there is a vague information for the response rates of the surrogate endpoints.

The mean parameters $\theta_{k,1}, \theta_{k,2}$ have an inverse gamma prior. Remember in section 4.3.3 we mentioned that we will put the inverse gamma priors on $\theta_{k,2}$ and Δ_k where $\Delta_k = \theta_{k,1} - \theta_{k,2}$, such that $\theta_{k,2} \sim \text{IG}(\alpha_{k,2}, \beta_{k,2}), \Delta_k \sim \text{IG}(\alpha_{k,0}, \beta_{k,0})$, to solve the label switching problem. In this simulation study, we assume that the expectations of $\theta_{k,2}$ and Δ_k equal to their theoretical values for treatment B (control group), such that $\theta_{B,2} = \frac{\beta_{k,2}}{\alpha_{k,2}-1}$, and $\Delta_B = \frac{\beta_{k,0}}{\alpha_{k,0}-1}$. So for scenarios Ia-If, IIa-IIc, and IIIa, we assume $\theta_{k,2} \sim \text{IG}(11, 70)$ and $\Delta_k \sim \text{IG}(11, 280)$; for scenarios IId-IIe, and IIIb, we assume $\theta_{k,2} \sim \text{IG}(11, 70)$ and $\Delta_k \sim \text{IG}(11, 170)$; and for scenarios IIIc, we assume $\theta_{k,2} \sim \text{IG}(11, 90)$ and $\Delta_k \sim \text{IG}(11, 490)$. Under this parameterization, we assume that the amount of information in these prior distributions is approximately equal to that from 11 subjects. Furthermore, we assume that the information is from

the historical data (or control group). This prior setup was chosen based on the recommendation of Huang et al. (2009), where they explained their considerations of choosing such a prior: the priors should be reasonably informative in order to show the difference between response categories of surrogate endpoint, yet they are not so strong that they can be altered by the data in the ongoing trial [Huang et al., 2009]. Normally, under the inverse gamma prior distribution, for example IG(α, β), if the investigators choose $\alpha/(\alpha+\beta)$ to be small (say, less than or equal to 0.1), then they wish to have a low prior weight on the historical data. And if they choose $\alpha/(\alpha+\beta)$ to be greater than 0.5, then they want to have a large prior weight on the historical data [Ibrahim et al., 2005]. In our simulation study, $\alpha_{k,0}/(\alpha_{k,0}+\beta_{k,0})$ for all the scenarios are smaller than 0.1, and $\alpha_{k,2}/(\alpha_{k,2}+\beta_{k,2})$ is slightly larger than 0.1 (11/(11+70)=0.13, 11/(11+90)=0.11). These parameterizations reflect we put a relatively low prior weight on the historical data.

As we mentioned before, the proposed algorithm will be compared with the standard response adaptive randomization, which does not use the information from the surrogate endpoint. For the standard RAR, we assume that the survival times for subjects in each treatment group have exponential distributions with mean parameters θ_A and θ_B, respectively. To estimate these unknown parameters, a Bayesian approach will be used as well. Similar as what we did for our proposed algorithm, we will put a vaguely informative prior on θ_A and θ_B, such that both θ_A and θ_B have an inverse gamma prior as IG(α_0, β_0), and $\beta_0/(\alpha_0-1)=\theta_B$. So for scenarios Ia-If, we assume $\theta_A, \theta_B \sim$ IG$(11, 182)$; for scenarios IIa-Ic, we assume $\theta_A, \theta_B \sim$ IG$(11, 294)$; for scenarios IId-Ie, we assume $\theta_A, \theta_B \sim$ IG$(11, 206)$; for scenario IIIa, we assume

$\theta_A, \theta_B \sim \text{IG}(11, 121)$; for scenario IIb, we assume $\theta_A, \theta_B \sim \text{IG}(11, 154)$; and for scenario IIIc, we assume $\theta_A, \theta_B \sim \text{IG}(11, 237)$. We will use the posterior means of θ_A and θ_B to obtain the target allocation proportion for the standard RAR procedure.

For the DBCD procedure, $\alpha = 2$ was chosen based on the recommendation of Rosenberger and Hu (2004), who showed that such a choice of α provides a reasonable trade-off between randomness and optimality [Rosenberger and Hu, 2004]. For each randomization procedure, 10,000 simulations per scenario was performed to evaluate the performance of our proposed algorithm. To assess the Markov Chain convergence, 3 chains with different set of initial values were used to test whether they all converge to the same target distribution.

4.5.4 Results

To investigate the performance of our proposed response adaptive randomization procedure (proposed RAR), we will compare the power, the observed allocation proportion to treatment A, the average number of subjects in treatment A, the average number of event, and the average total observed survival time under the proposed algorithm with those under the equal allocation randomization and standard response adaptive randomization.

Table 4.2 shows the comparison of the three randomization procedures in terms of power for the Wald test. Except for scenarios Ie, If, and IIIc, both the standard response adaptive randomization and our proposed adaptive randomization procedure yield a larger power compared to equal allocation randomization. And for scenarios Ie, If, and IIIc, the power under the proposed RAR algorithm is similar to that under

the equal allocation randomization procedure. This is consistent with the conclusion made by Zhang and Rosenberger that the power under the double-adaptive biased coin design was as powerful or slightly more powerful than the equal randomization procedure [Rosenberger and Hu, 2004]. Moreover, the power under our proposed RAR algorithm is comparable as that under the standard RAR procedure.

Table 4.3 shows the average number of subjects assigned to treatment A, along with their standard deviations. From table 4.3, one can see that both the standard response adaptive randomization and our proposed response adaptive randomization are more ethical than the equal allocation randomization procedure, since both of them assigned more subjects to treatment A (the group that has a larger mean survival time), as compared to the equal allocation. Meanwhile, except for scenarios IId and IIe where the number of subjects assigned to treatment A are the same for standard and proposed RAR, our proposed algorithm tends to allocate slightly more subjects to treatment A as compared to the standard RAR. On average, our proposed algorithm assigns 1 to 5 more subjects to the better treatment group, compared with the standard RAR procedure.

Table 4.4 shows the observed allocation proportion to treatment A, along with their standard deviations. As expected, both the standard response adaptive randomization and the proposed adaptive randomization have an allocation proportion to treatment A greater then 0.5. The allocation proportion under the proposed randomization procedure is slightly greater than that under the standard RAR for all the experimental scenarios. However, it can be seen from table 4.4 that our proposed algorithm has higher standard deviations of allocation proportions than the standard

RAR procedure. When there is a larger percentage of censoring, the difference of observed allocation proportion between the standard and proposed RAR is larger as well. For instance, when the proportion of censoring decreases (as in scenarios Ia-If where the proportion of non-censoring in treatment A increases from 0.49 to 0.70), the difference of observed allocation proportion to treatment A decreases from 0.024 to 0.002. Thus, the benefit of assigning more subjects to a better treatment under our proposed algorithm is more obvious when there is a relatively large delay in the primary outcome, as compared to the standard RAR procedure.

Table 4.5 and table 4.6 show the average number of events and the average total observed survival time for each of the three randomization procedure, along with the standard deviations, respectively. Both the standard response adaptive randomization and our proposed adaptive randomization reduce the average number of events (deaths) as compared to the equal allocation procedure. For all the experimental scenarios, 3-6 fewer events for both the standard RAR and proposed RAR as compared to the equal allocation randomization. Compared with standard RAR procedure, our proposed algorithm has a comparable or a slightly reduced number of events. On average, our proposed algorithm has 1 fewer event than the standard RAR procedure for scenarios Ib, Ic, IIIa, and IIIc, and a same number of events as standard RAR for other scenarios. Even through these reductions is not huge, we still think any reduction in the number of events is desirable in survival trials. Also, it is known that in survival analysis, power is directly related to the number of events. Therefore, one cannot expect a large reduction in the number of events without sacrificing the power [Sverdlov et al., 2011]. In table 4.6, one can see that except for scenarios IIa-IIc, both

the standard and proposed RAR have a relatively higher observed total survival time than the equal allocation randomization procedure. For scenarios IIa-IIc, the equal allocation procedure has a larger total observed survival time. That may because for these three scenarios, the percentage of censoring is pretty large, thus at the end of the trial, the number of subjects having an event that can be used in the analysis is small. Our proposed algorithm remains a comparable total observed survival time as the standard response adaptive randomization procedure.

4.6 Discussion

In this chapter, we proposed a new response adaptive randomization procedure for clinical trials with survival primary endpoints. In clinical trials, especially for survival trials, censoring is very common. At the meantime, surrogate endpoints can always be obtained sooner than the primary survival time. When the surrogate endpoints becomes available, we should not ignore those information. Thus, under our proposed algorithm, we connect the surrogate endpoint with the primary survival time, and use these information in the adaptive randomization procedure. Specifically, we model the relationship between the surrogate and primary endpoint through a mixture model (a mixture of exponential distributions), and estimate the parameters of interest through a Bayesian approach. Through simulation studies, we find that our proposed response adaptive randomization is more effective in assigning subjects to better treatments as compared with equal allocation randomization and as effective as or even performs better than the standard response adaptive randomization

procedure. Specifically, the proposed algorithm tends to allocate more subjects to the better performance group, decreases the average number of events, and remains a comparable power as compared to the standard response adaptive randomization procedure.

Table 4.1: Parameter setup for different experimental scenario

Scenario	Sample Size	Treatment A			Treatment B		
		$(p_{1,A}, p_{2,A})$	$(\theta_{1,A}, \theta_{2,A})$	ϵ_A	$(p_{1,B}, p_{2,B})$	$(\theta_{1,B}, \theta_{2,B})$	ϵ_B
Ia	66	(0.7, 0.3)	(76, 9)	0.49	(0.4, 0.6)	(35, 7)	0.80
Ib	74	(0.7, 0.3)	(68, 9)	0.53	(0.4, 0.6)	(35, 7)	0.80
Ic	90	(0.7, 0.3)	(58, 9)	0.57	(0.4, 0.6)	(35, 7)	0.80
Id	134	(0.7, 0.3)	(47, 9)	0.63	(0.4, 0.6)	(35, 7)	0.80
Ie	182	(0.7, 0.3)	(41, 8)	0.67	(0.4, 0.6)	(35, 7)	0.80
If	246	(0.7, 0.3)	(36, 8)	0.70	(0.4, 0.6)	(35, 7)	0.80
IIa	150	(0.8, 0.2)	(76, 9)	0.46	(0.8, 0.2)	(35, 7)	0.69
IIb	192	(0.8, 0.2)	(68, 9)	0.49	(0.8, 0.2)	(35, 7)	0.69
IIc	302	(0.8, 0.2)	(58, 9)	0.54	(0.8, 0.2)	(35, 7)	0.69
IId	142	(0.8, 0.2)	(48, 9)	0.60	(0.8, 0.2)	(24, 7)	0.78
IIe	224	(0.8, 0.2)	(41, 9)	0.64	(0.8, 0.2)	(24, 7)	0.78
IIIa	132	(0.7, 0.3)	(58, 9)	0.57	(0.3, 0.7)	(58, 9)	0.74
IIIb	158	(0.7, 0.3)	(35, 7)	0.71	(0.3, 0.7)	(35, 7)	0.83
IIIc	226	(0.7, 0.3)	(24, 7)	0.80	(0.3, 0.7)	(24, 7)	0.97

Table 4.2: Summary of power of the three randomization procedures using the two-sided Wald test

Scenario	Sample size (N)	Equal Allocation	Standard RAR	Proposed RAR
Ia	66	0.840	0.876	0.875
Ib	74	0.844	0.867	0.865
Ic	90	0.846	0.871	0.864
Id	134	0.871	0.881	0.871
Ie	182	0.862	0.869	0.861
If	246	0.843	0.844	0.842
IIa	150	0.826	0.844	0.844
IIb	192	0.832	0.846	0.846
IIc	302	0.832	0.837	0.840
IId	142	0.824	0.842	0.847
IIe	224	0.834	0.845	0.837
IIIa	132	0.849	0.849	0.860
IIIb	158	0.847	0.855	0.855
IIIc	226	0.849	0.848	0.846

Table 4.3: Summary of average number of subjects assigned to treatment A, with standard deviations (SD) for the three randomization procedures

Scenario	Sample size (N)	Equal Allocation	Standard RAR	Proposed RAR
Ia	66	33 (4.05)	47 (4.63)	49 (5.12)
Ib	74	37 (4.30)	53 (5.21)	55 (5.63)
Ic	90	45 (4.78)	64 (6.24)	66 (6.78)
Id	134	67 (5.85)	94 (8.94)	96 (9.53)
Ie	182	91 (6.74)	124 (11.53)	126 (12.20)
If	246	123 (7.85)	162 (14.49)	163 (15.42)
IIa	150	75 (6.14)	102 (10.53)	104 (10.72)
IIb	192	96 (6.93)	128 (12.67)	131 (13.30)
IIc	302	151 (8.79)	195 (18.15)	199 (18.91)
IId	142	71 (5.99)	96 (9.68)	96 (9.56)
IIe	224	112 (7.51)	147 (13.68)	147 (13.53)
IIIa	132	66 (5.78)	91 (8.59)	96 (9.90)
IIIb	158	79 (6.30)	108 (10.36)	112 (10.47)
IIIc	226	113 (7.53)	149 (13.38)	151 (12.47)

Table 4.4: Summary of observed allocation proportion to treatment A, with standard deviations (SD) for the three randomization procedures

Scenario	Sample size (N)	Equal Allocation	Standard RAR	Proposed RAR
Ia	66	0.500 (0.0613)	0.715 (0.0702)	0.739 (0.0775)
Ib	74	0.500 (0.0581)	0.713 (0.0704)	0.738 (0.0760)
Ic	90	0.500 (0.0531)	0.709 (0.0693)	0.733 (0.0754)
Id	134	0.500 (0.0437)	0.699 (0.0667)	0.715 (0.0711)
Ie	182	0.500 (0.0370)	0.679 (0.0634)	0.690 (0.0670)
If	246	0.500 (0.0319)	0.660 (0.0589)	0.662 (0.0627)
IIa	150	0.500 (0.0409)	0.679 (0.0702)	0.693 (0.0714)
IIb	192	0.500 (0.0361)	0.668 (0.0660)	0.683 (0.0693)
IIc	302	0.500 (0.0291)	0.646 (0.0601)	0.660 (0.0626)
IId	142	0.500 (0.0422)	0.678 (0.0681)	0.676 (0.0673)
IIe	224	0.500 (0.0335)	0.656 (0.0611)	0.656 (0.0604)
IIIa	132	0.500 (0.0438)	0.691 (0.0650)	0.731 (0.0750)
IIIb	158	0.500 (0.0399)	0.684 (0.0656)	0.706 (0.0663)
IIIc	226	0.500 (0.0333)	0.658 (0.0592)	0.668 (0.0552)

Table 4.5: Summary of average number of events, with standard deviations (SD) for the three randomization procedures

Scenario	Sample size (N)	Equal Allocation	Standard RAR	Proposed RAR
Ia	66	46 (3.73)	42 (4.10)	42 (4.15)
Ib	74	52 (3.87)	49 (4.36)	48 (4.34)
Ic	90	65 (4.22)	62 (4.70)	61 (4.77)
Id	134	100 (5.03)	96 (5.47)	96 (5.55)
Ie	182	139 (5.71)	135 (6.21)	135 (6.12)
If	246	191 (6.55)	187 (6.93)	187 (7.02)
IIa	150	93 (5.93)	88 (6.54)	88 (6.67)
IIb	192	121 (6.63)	116 (7.31)	116 (7.34)
IIc	302	197 (8.32)	191 (8.91)	191 (8.91)
IId	142	101 (5.36)	97 (9.68)	97 (5.91)
IIe	224	164 (6.68)	160 (7.08)	160 (7.11)
IIIa	132	94 (5.19)	90 (5.62)	89 (5.78)
IIIb	158	125 (5.08)	122 (5.49)	122 (5.45)
IIIc	226	190 (5.47)	188 (5.81)	187 (5.83)

Table 4.6: Summary of total observed survival time, with standard deviations (SD) for the three randomization procedures

Scenario	Sample size (N)	Equal Allocation	Standard RAR	Proposed RAR
Ia	66	627 (117.75)	642 (123.25)	640 (122.61)
Ib	74	717 (124.02)	738 (133.29)	739 (130.56)
Ic	90	892 (138.67)	926 (145.05)	932 (146.02)
Id	134	1358 (167.82)	1420 (176.44)	1427 (178.78)
Ie	182	1842 (192.62)	1917 (199.46)	1925 (201.80)
If	246	2494 (220.70)	2598 (231.26)	2595 (228.77)
IIa	150	1676 (203.16)	1635 (206.53)	1628 (208.42)
IIb	192	2182 (229.01)	2155 (237.70)	2142 (232.77)
IIc	302	3513 (287.40)	3486 (294.87)	3485 (291.87)
IId	142	1645 (185.74)	1660 (190.01)	1657 (190.63)
IIe	224	2623 (232.30)	2648 (233.43)	2654 (234.64)
IIIa	132	1300 (166.88)	1353 (176.52)	1363 (177.98)
IIIb	158	1527 (169.04)	1615 (181.30)	1627 (178.89)
IIIc	226	2105 (181.25)	2201 (193.09)	2209 (191.32)

Chapter 5

Discussion and Future Work

The standard response adaptive randomization procedure will be affected by the delayed primary endpoint. When there is a large delay time in the primary endpoint, less information can be used in the randomization procedure, thus the benefit of assigning more patients to a better treatment group decreases. A surrogate endpoint is a measurement made after the treatment to determine whether the treatment is working. Normally, a surrogate endpoint can be obtained earlier than the primary endpoint. When surrogate endpoint becomes available, the information from the surrogate endpoint is valuable and should not be ignored. Therefore, we proposed two response adaptive randomization procedures which will connect the surrogate endpoint with the primary endpoint through a statistical model and use the accumulated information from both the surrogate and primary endpoints to skew the allocation proportion.

We first proposed a response adaptive randomization for clinical trials with nor-

mally distributed primary and surrogate endpoints. We connected the surrogate and primary endpoints through a bivariate normal distribution model, and introduced a Bayesian approach to estimate the unknown parameters. We then substituted the mean and standard deviation of primary endpoint in the target allocation function with the conditional mean and conditional variance of primary endpoint. As anticipated, the proposed algorithm can assign more patients to the more efficacious treatment group, and is more powerful than the traditional equal allocation randomization procedure. In addition, our proposed algorithm has advantages over the standard response adaptive randomization which is solely based on the information from the primary endpoint. Compared to the standard RAR, our proposed response adaptive randomization can allocate slightly more patients to the superior treatment group, and is more robust when there is a large delay in the primary endpoint.

We then consider a clinical trial with survival responses. We proposed a response adaptive randomization for clinical trials with survival primary endpoint and categorical surrogate endpoint. We modeled the relationship between the survival time and surrogate endpoint through a mixture model, and obtained the posterior means of parameters of interest through a Bayesian model. We then calculated the target allocation ratio using the posterior estimates of marginal mean survival times. Through some simulations, we found that the proposed response adaptive can allocate more patients to a better performance treatment group, decrease the average number of events (deaths), and maintain a comparable power as compared to the traditional equal allocation randomization procedure. Meanwhile, compared to the standard response adaptive randomization where only primary endpoint is used in

the allocation procedure, our proposed algorithm is as effective as, and in most cases, it performs even better in assigning subjects to better treatments.

There are some limitations with our proposed algorithm. First, both the two proposed approaches assume that the surrogate endpoint can be observed before the next patient coming in the trial, which is not always the case. In really clinical trials, most of the surrogate endpoints cannot be observed immediately after the treatment, and there is usually a lag time to observe a surrogate endpoint. Normally, the surrogate endpoint can be obtained much sooner than the primary endpoint. Therefore, it will be more appropriate to model the delay in both the primary and surrogate endpoint. For example, we can assume that the delay time has an exponential distribution, and the mean delay time in the surrogate endpoint is less than that in the primary endpoint. Second, our proposed algorithm only considered one surrogate endpoint. And for normally distributed primary endpoint, we assume the surrogate endpoint also has a normal distribution; for survival responses, we assume that the surrogate endpoint has a multinomial distribution. However, one may have more than one surrogate endpoint available, and these surrogate endpoints may have distributions other than normal and multinomial. Thus, in the future, we can consider more than one surrogate endpoint with some other distributions. Third, in chapter 3, we considered a clinical trial with normally distributed primary and surrogate endpoints. We investigated the performance of our proposed algorithm under different correlations between the surrogate and primary endpoint, and found that the correlation does not have an impact on the performance. As we know, the correlation measures the linear relationship between the two variables. Instead of using the correlation,

we can consider the concordance between the surrogate and primary endpoint in the randomization procedure. Different from the correlation, the concordance measures the agreement between two measurements. A strong linear correlation between two variables does not mean there is also a strong concordance. Therefore, we may investigate the performance of our proposed algorithm under different concordance in the future. Fourth, in the simulation study for both normally and exponentially distributed primary outcomes, we only looked at a small number of scenarios. In the future, we may want to change the parameter setups so we can evaluate the performance of our proposed algorithm under more scenarios. Fifth, in the doubly adaptive biased coin design (DBCD), we used $\alpha = 2$ as recommonded by Zhang and Rosenberger. For the future works, we may want to look at some other choices of α in the DBCD function. Finally, in the response adaptive randomization procedure, the allocation of the next patient dependents on the performance of the previous patients, and this may have a violation of the assumption that the patients in the study are independent.

Bibliography

Administration, U. F. . D. et al. (2010). Guidance for industry–adaptive design clinical trials for drugs and biologics (draft guidance).

Atkinson, A. C. and Biswas, A. (2013). *Randomised response-adaptive designs in clinical trials*. CRC Press.

Biswas, A., Bhattachary, R., and Zhang, L. (2007). Optimal response-adaptive designs for continuous responses in phase iii trials. *Biometrical Journal*, 49(6):928–940.

Biswas, A. and Bhattacharya, R. (2012). Response-adaptive designs for continuous treatment responses in phase iii clinical trials: A review. *Statistical methods in medical research*, page 0962280212441424.

Biswas, A. and Mandal, S. (2004). Optimal adaptive designs in phase iii clinical trials for continuous responses with covariates. In *mODa 7Advances in Model-Oriented Design and Analysis*, pages 51–59. Springer.

Casella, G., Mengersen, K., Robert, C., and Titterington, D. (2002). Perfect slice

samplers for mixtures of distributions. *Journal of the Royal Statistical Society: Series B*, 64(4):777790.

Chow, S.-C., Chang, M., et al. (2008). Adaptive design methods in clinical trials-a review. *Orphanet J Rare Dis*, 3(11):169–90.

Deb, P. et al. (2008). Finite mixture models.

Dreze, J. H. and Richard, J.-F. (1983). Handbook of econometrics, volume i, chapter bayesian analysis of simultaneous equation systems.

Eisele, J. R. and Woodroofe, M. B. (1995). Central limit theorems for doubly adaptive biased coin designs. *The Annals of Statistics*, pages 234–254.

Escoto, B. (2013). Bayesian claim severity with mixed distributions. *Variance*, 7(2):110–122.

Food, Administration, D., et al. (2007). Critical path opportunities list. 2006.

Fountzilas, G., Ciuleanu, E., Dafni, U., Plataniotis, G., Kalogera-Fountzila, A., Samantas, E., Athanassiou, E., Tzitzikas, J., Ciuleanu, T., Nikolaou, A., et al. (2004). Concomitant radiochemotherapy vs radiotherapy alone in patients with head and neck cancer. *Medical Oncology*, 21(2):95–107.

Fu, X., Shen, Z., Chen, Y., Xie, J., Guo, Z., Zhang, M., and Sheng, Z. (1998). Randomised placebo-controlled trial of use of topical recombinant bovine basic fibroblast growth factor for second-degree burns. *The lancet*, 352(9141):1661–1664.

Gallo, P., Chuang-Stein, C., Dragalin, V., Gaydos, B., Krams, M., and Pinheiro, J. (2006). Adaptive designs in clinical drug developmentan executive summary of the phrma working group. *Journal of biopharmaceutical statistics*, 16(3):275–283.

Gelman, A., Shirley, K., et al. (2011). Inference from simulations and monitoring convergence. *Handbook of Markov chain Monte Carlo*, pages 163–174.

Group, B. D. W. (2001). Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clinical Pharmacology and Therapeutics*, 69(3):89–95.

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

Hu, F., Zhang, L.-X., Cheung, S. H., and Chan, W. S. (2008). Doubly adaptive biased coin designs with delayed responses. *Canadian Journal of Statistics*, 36(4):541–559.

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

Ibrahim, J. G., Chen, M.-H., and Sinha, D. (2005). *Bayesian survival analysis*. Wiley Online Library.

Mahajan, R. and Gupta, K. (2010). Adaptive design clinical trials: Methodology, challenges and prospect. *Indian journal of pharmacology*, 42(4):201.

Marin, J.-M., Mengersen, K., and Robert, C. P. (2005). Bayesian modelling and inference on mixtures of distributions. *Handbook of statistics*, 25:459–507.

McLachlan, G. and Peel, D. (2004). *Finite mixture models*. John Wiley & Sons.

Murphy, K. P. (2007). Conjugate bayesian analysis of the gaussian distribution. *def*, 1(2σ2):16.

Nowacki, A. S., Zhao, W., and Palesch, Y. Y. (2015). A surrogate-primary replacement algorithm for response-adaptive randomization in stroke clinical trials. *Statistical methods in medical research*, page 0962280214567142.

Plummer, M. et al. (2003). Jags: A program for analysis of bayesian graphical models using gibbs sampling. In *Proceedings of the 3rd international workshop on distributed statistical computing*, volume 124, page 125. Vienna.

Rosenberger, W. F. et al. (1993). Asymptotic inference with response-adaptive treatment allocation designs. *The Annals of Statistics*, 21(4):2098–2107.

Rosenberger, W. F. and Hu, F. (2004). Maximizing power and minimizing treatment failures in clinical trials. *Clinical Trials*, 1(2):141–147.

Rosenberger, W. F. and Lachin, J. M. (2004). *Randomization in clinical trials: theory and practice*. John Wiley & Sons.

Rosenberger, W. F. and Seshaiyer, P. (1997). Adaptive survival trials. *Journal of Biopharmaceutical Statistics*, 7(4):617–624.

Shi, Q. (2007). *Bayesian Methods of Evaluation and Use of Surrogate Endpoints in Single-trial Settings*. ProQuest.

Sinks, S. (2013). Response adaptive design using auxiliary and primary outcomes.

Sverdlov, O., Tymofyeyev, Y., and Wong, W. K. (2011). Optimal response-adaptive randomized designs for multi-armed survival trials. *Statistics in medicine*, 30(24):2890–2910.

Wilson, T. W., Lacourciere, Y., Barnes, C. C., et al. (1998). The antihypertensive efficacy of losartan and amlodipine assessed with office and ambulatory blood pressure monitoring. *Canadian Medical Association Journal*, 159(5):469–476.

Xu, J. and Yin, G. (2014). Two-stage adaptive randomization for delayed response in clinical trials. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 63(4):559–578.

Yin, G. (2013). *Clinical trial design: Bayesian and frequentist adaptive methods*, volume 876. John Wiley & Sons.

Zhang, L. and Rosenberger, W. F. (2006). Response-adaptive randomization for clinical trials with continuous outcomes. *Biometrics*, 62(2):562–569.

Zhang, L. and Rosenberger, W. F. (2007). Response-adaptive randomization for survival trials: the parametric approach. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 56(2):153–165.

Appendix A

Chapter 3 appendix

A.1 Bivariate normal distribution model in JAGS

```
1
2  model{
3
4    # set values for the parameters in the prior distribution
5    R[1,1] <- 1
6    R[1,2] <- 0
7    R[2,1] <- 0
8    R[2,2] <- 1
9    nu <- 4
10   mu0[1] <- 0
11   mu0[2] <- 0
12   lambda <- 0.001
13
14   #-------- likelihood ------------#
15
16   for (i in 1:N) {
17
18     # marginal distribution of the surrogate endpoint
```

```
    Sdata[i] ~ dnorm(mu[1], taus)

    # conditional mean of primary given surrogate
    muTgvS[i] <- mu[2] + T*(Sdata[i] - mu[1])
    # conditional distribution of primary endpoint given surrogate
    Tdata[i] ~ dnorm(muTgvS[i], tauTgvS)
  }

  # Gamma prior distribution on marginal precision of S
  as <- (nu-1)/2
  bs <- R[1,1]/2
  taus ~ dgamma(as, bs)

  # Gamma prior distribution on conditional precision of Primary given surrogate
  at <- nu/2
  bt1 <- R[2,2] - (R[2,1])^2/R[1,1]
  bt <- bt1/2
  tauTgvS ~ dgamma(at, bt)

  # Normal prior distribution on Omega_st, where Omega_st is a scalar
  prmean <- R[1,2] / R[1,1]
  prprec <- R[1,1] * tauTgvS
  T ~ dnorm(prmean, prprec)

  #---------- transformations to quantities of interest ------------#
  # variance of S
  Sigma[1,1] <- 1/taus
  # Covariance between S and T
  Sigma[1,2] <- T*Sigma[1,1]
  # Variance of T
  Sigma[2,2] <- 1/tauTgvS + (Sigma[1,2])^2/Sigma[1,1]
  Sigma[2,1] <- Sigma[1,2]

  # normal prior for mean vectors
   tau.mu <- inverse(Sigma)*lambda
```

```
    mu ~ dmnorm(mu0, tau.mu)

}
```

A3_1.R

A.2 Proposed response adaptive randomization procedure for normally distributed primary outcome

```
continuous.randomization<-function(isim.start, # start from zero
                                   seed, # used in sed.seed function
                                   nSims, # number of simulation
                    nSbjs, # total number of patients in the trial
                    n.init, # number of subjects in the equal randomizaition
                    alpha, # nonnegative number in DBCD procedure
                    Chi_square, # Chi-square for the hypothesis test
                    lambda1, # delay time of primary endpoint for treatment A
                    lambda2, # delay time of primary endpoint for treatment B
                    lambda3, # entry time for both treatment groups
                    mu.SA, mu.SB, mu.TA, mu.TB,
                    sigma.SSA, sigma.SSB, sigma.TTA, sigma.TTB, corr) {

  allocation.prob.matrix = matrix(, nrow=nSims-isim.start, ncol=nSbjs)

  power = {}
  expected.response = {}
  N.A.final.vec = {}
  N.B.final.vec = {}
  allocation.prop.final.vec = {}
  TS.vec = {}
```

```
convergence.A.vec = {}
convergence.B.vec = {}

mu.A <- c(mu.SA, mu.TA)  # mean vector that used to simulate surrogate and primary
     endpoint for treatment A
mu.B <- c(mu.SB, mu.TB)

sigma.STA <- corr*sqrt(sigma.SSA)*sqrt(sigma.TTA)
sigma.STB <- corr*sqrt(sigma.SSB)*sqrt(sigma.TTB)
Sigma.A <- matrix(c(sigma.SSA, sigma.STA, sigma.STA, sigma.TTA), 2, 2) #
    Covariance matrix for treatment A
Sigma.B <- matrix(c(sigma.SSB, sigma.STB, sigma.STB, sigma.TTB), 2, 2)

##################################################
################## DBCD procedure ################
##################################################

DBCD <- function(muA, sigmaA, muB, sigmaB, alpha, N.A, N.B) {
  if (is.na(muA) | is.na(muB) | is.na(sigmaA) | is.na(sigmaB) | muA < 0 | muB < 0)
       {
    g.function = 0.5
  } else {
    r = sigmaA*sqrt(muB)/sigmaB/sqrt(muA)
    s = ifelse(muA < muB & r >1 | muA > muB & r<1, 1, 0)
    rho = ifelse(s==1, sigmaA*sqrt(muB)/(sigmaA*sqrt(muB)+sigmaB*sqrt(muA)), 0.5)
    x <- N.A/(N.A+N.B)
    y <- rho
    if (x==0) {
      g.function = 1
    } else if (x==1) {
      g.function = 0
    } else{
      g.function = y*(y/x)^alpha/(y*(y/x)^alpha+(1-y)*((1-y)/(1-x))^alpha)
    }
  }
```

```
54    return(g.function)
55  }
56
57
58  set.seed(seed+isim.start)
59
60  for (isim in (isim.start+1):nSims) {
61
62    S.data = {}       # surrogate endpoint
63    T.data = {}       # primary endpoint
64    trt = {}          # treatment indicator
65    et = {} # entry time
66    et.interval = {}
67    dt.interval = {}
68    dt = {} # observed time
69    #delta = {} # indicator variable that equals to 1 if the primary endpoint is
          observed
70
71    convergence.A = {}
72    convergence.B = {}
73
74    ##### first, assign n.init subjects to each treatment, and assume that the 2*n.
          init subjects' primary outcome and surrogate outcome are available
          immediately.
75
76    for (i.equal in 1:(2*n.init)) {
77      # first, record the patient's entry time
78      entry.t <- rexp(1,1/lambda3)
79      et.interval <- rbind(et.interval, entry.t)
80
81      U = runif(1, min=0, max=1)
82      if(U < 0.5) {
83        A.init = mvrnorm(1, mu.A, Sigma.A)
84        S.data = rbind(S.data, A.init[1])
85        T.data = rbind(T.data, A.init[2])
```

```
        trt = rbind(trt, 1)

        # simulate the delay time for this patient
        delay.t <- rexp(1, 1/lambda1)
        dt.interval <- rbind(dt.interval, delay.t)

        et = rbind(et, 0)
        dt = rbind(dt, 0)
        #delta = rbind(delta, 1)
      } else {
        B.init = mvrnorm(1, mu.B, Sigma.B)
        S.data = rbind(S.data, B.init[1])
        T.data = rbind(T.data, B.init[2])
        trt = rbind(trt, 2)

        # simulate the delay time for this patient
        delay.t <- rexp(1, 1/lambda2)
        dt.interval <- rbind(dt.interval, delay.t)

        et = rbind(et, 0)
        dt = rbind(dt, 0)
        #delta = rbind(delta, 1)
      }
    }

    ## entry time and observed time
    et[1] = et.interval[1]
    dt[1] = et[1] + dt.interval[1]
    for (i in 2:(2*n.init)) {
      et[i] = et[i-1] + et.interval[i]
      dt[i] = et[i] +dt.interval[i]
    }

    #### randomization procedure
    for (iSbj in (2*n.init+1):nSbjs) {
```

```
121
122       # first, record the patient's entry time
123       entry.t <- rexp(1,1/lambda3)
124       et.interval <- rbind(et.interval, entry.t)
125       et[iSbj] <- et[iSbj-1]+entry.t
126
127       # then we need to update the indicator variable delta before we assign this
              patient
128       threshold <- et[iSbj]
129       delta <- ifelse(dt<=threshold, 1, 0)
130
131       ##############################################################################
132       ## For treatment A and B, get the posterior means for each parameters ####
133       ##############################################################################
134       SA.data <- S.data[trt==1]
135       TA.data <- T.data[trt==1]
136       delta.A <- delta[trt==1]
137       TA.data[delta.A==0] <- NA
138       gibbs.data.A <- list(Sdata=SA.data, Tdata=TA.data, N=length(SA.data))
139       inits.A <- list(list(mu=c(0,0), taus=1, tauTgvS=1, T=0, .RNG.seed=12345,
140                            .RNG.name="base::Mersenne-Twister"),
141                                 list(mu=c(2,1), taus=0.5, tauTgvS=0.5, T=0.5, .RNG.seed
                                        =123456,
142                            .RNG.name="base::Mersenne-Twister"),
143                                 list(mu=c(1,2), taus=2, tauTgvS=1.5, T=1, .RNG.seed
                                        =1234,
144                            .RNG.name="base::Mersenne-Twister"))
145             gibbs.jags.A <- jags.model(file="gibbs_wishart.bug", n.chains=3, inits=
                    inits.A, data=gibbs.data.A)
146             gibbs.out.A <- coda.samples(gibbs.jags.A, n.iter=5E3, thin=2, variable.
                    names=c("mu", "taus", "tauTgvS", "T"))
147             gelman.A <- gelman.diag(gibbs.out.A, multivariate=FALSE)
148             convergence.A.psrf <- ifelse(gelman.A$psrf[,1]<1.1, 0, 1)
149             convergence.A.ind <- sum(convergence.A.psrf)
150             convergence.A <- rbind(convergence.A, convergence.A.ind)
```

```
151             out.A <- summary(gibbs.out.A)
152             T.A <- out.A$statistics[1,1]
153             sigmaTgvS.A <- 1/out.A$statistics[4,1]
154             sigmaTgvS.A.sqrt <- sqrt(sigmaTgvS.A)
155             muT.A <- out.A$statistics[3,1]
156             muS.A <- out.A$statistics[2,1]
157             muTgvS.A <- muT.A + T.A*(mean(SA.data)-muS.A)
158             N.A <- length(SA.data)
159
160             SB.data <- S.data[trt==2]
161         TB.data <- T.data[trt==2]
162         delta.B <- delta[trt==2]
163         TB.data[delta.B==0] <- NA
164         gibbs.data.B <- list(Sdata=SB.data, Tdata=TB.data, N=length(SB.data))
165         inits.B <- list(list(mu=c(0,0), taus=1, tauTgvS=1, T=0, .RNG.seed=12345,
166                         .RNG.name="base::Mersenne-Twister"),
167                             list(mu=c(2,1), taus=0.5, tauTgvS=0.5, T=0.5, .RNG.seed
                                    =123456,
168                         .RNG.name="base::Mersenne-Twister"),
169                             list(mu=c(1,2), taus=2, tauTgvS=1.5, T=1, .RNG.seed
                                    =1234,
170                         .RNG.name="base::Mersenne-Twister"))
171             gibbs.jags.B <- jags.model(file="gibbs_wishart.bug", n.chains=3, inits=
                    inits.B, data=gibbs.data.B)
172             gibbs.out.B <- coda.samples(gibbs.jags.B, n.iter=5E3, thin=2, variable.
                    names=c("mu", "taus", "tauTgvS", "T"))
173             gelman.B <- gelman.diag(gibbs.out.B, multivariate=FALSE)
174             convergence.B.psrf <- ifelse(gelman.B$psrf[,1]<1.1, 0, 1)
175             convergence.B.ind <- sum(convergence.B.psrf)
176             convergence.B <- rbind(convergence.B, convergence.B.ind)
177             out.B <- summary(gibbs.out.B)
178             T.B <- out.B$statistics[1,1]
179             sigmaTgvS.B <- 1/out.B$statistics[4,1]
180             sigmaTgvS.B.sqrt <- sqrt(sigmaTgvS.B)
181             muT.B <- out.B$statistics[3,1]
```

```
182             muS.B <- out.B$statistics[2,1]
183             muTgvS.B <- muT.B + T.B*(mean(SB.data)-muS.B)
184             N.B <- length(SB.data)
185 
186             # calculate the allocation ratio, and then do the randomization process
187       U = runif(1,min=0, max=1)
188 
189       allocation.interim = DBCD(muA=muTgvS.A, sigmaA=sigmaTgvS.A.sqrt, muB=muTgvS.B,
              sigmaB=sigmaTgvS.B.sqrt, alpha=alpha, N.A=N.A, N.B=N.B)
190 
191       allocation.prob.matrix[isim-isim.start, iSbj] = allocation.interim
192 
193       if (U < allocation.interim) {
194         # allocate this patient to trt A
195         sim.A <- mvrnorm(n=1, mu=mu.A, Sigma=Sigma.A)
196         S.data = rbind(S.data, sim.A[1])
197         T.data = rbind(T.data, sim.A[2])
198         trt = rbind(trt, 1)
199 
200         # simulate the delay time for this patient
201         delay.t <- rexp(1, 1/lambda1)
202         dt.interval <- rbind(dt.interval, delay.t)
203         dt[iSbj] <- et[iSbj] + delay.t
204 
205       } else {
206         sim.B <- mvrnorm(n=1, mu=mu.B, Sigma=Sigma.B)
207         S.data = rbind(S.data, sim.B[1])
208         T.data = rbind(T.data, sim.B[2])
209         trt = rbind(trt, 2)
210 
211         # simulate the delay time for this patient
212         delay.t <- rexp(1, 1/lambda2)
213         dt.interval <- rbind(dt.interval, delay.t)
214         dt[iSbj] <- et[iSbj] + delay.t
215       }
```

```
    }

    ##########################################################################
    ##### point estimate of mean and variance ############################
    ##########################################################################
    data.final.A <- T.data[trt==1]
    data.final.B <- T.data[trt==2]
    N.A.final <- length(data.final.A)
    N.B.final <- length(data.final.B)
    mu.A.est <- mean(data.final.A)
    mu.B.est <- mean(data.final.B)
    sigma2.A.est <- var(data.final.A)
    sigma2.B.est <- var(data.final.B)

    TS = (mu.A.est - mu.B.est)^2/(sigma2.A.est/N.A.final + sigma2.B.est/N.B.final)
    TS.vec = rbind(TS.vec, TS)

    N.A.final.vec <- rbind(N.A.final.vec, N.A.final)
    N.B.final.vec <- rbind(N.B.final.vec, N.B.final)

    allocation.prop.final <- N.A.final/(N.A.final+N.B.final)
    allocation.prop.final.vec <- rbind(allocation.prop.final.vec, allocation.prop.
        final)

    sd.A.est <- sqrt(sigma2.A.est)
    sd.B.est <- sqrt(sigma2.B.est)

    total.response <- sum(T.data)
    #total.response <- (N.A.final+N.B.final)*(rho.hat*mu.A.est+(1-rho.hat)*mu.B.est)
    expected.response <- rbind(expected.response, total.response)

    ## convergence
    convergence.A.vec <- rbind(convergence.A.vec, sum(convergence.A))
    convergence.B.vec <- rbind(convergence.B.vec, sum(convergence.B))
}
```

```
250
251   power = ifelse(TS.vec > Chi_square, 1, 0)
252   power.est = sum(power)/length(power)
253   N.A.est = mean(N.A.final.vec)
254   N.B.est = mean(N.B.final.vec)
255   #rho.est = mean(rho.final)
256   total.est = mean(expected.response)
257
258   list(power.est=power.est, N.A.est=N.A.est, N.B.est=N.B.est, total.est=total.est,
259        N.A.final.vec = N.A.final.vec, N.B.final.vec = N.B.final.vec, allocation.prop
             .final.vec = allocation.prop.final.vec, expected.response = expected.
             response, allocation.prob.matrix=allocation.prob.matrix,
260        TS.vec = TS.vec, power=power, power.est=power.est, convergence.A.vec =
             convergence.A.vec, convergence.B.vec = convergence.B.vec)
261
262
263 }
```

A3_2.R

Appendix B

Chapter 4 appendix

B.1 Mixture model in JAGS

```
model {
  ## likelihood
  for (i in 1:N) {
    is.censored[i] ~ dinterval(t.to.event[i], t.cen[i])
    t.to.event[i] ~ dexp(lambda[i])
    lambda[i] <- lambda.vec[surrogate[i]]
    surrogate[i] ~ dcat(pi[1:2])
  }

  ## prior
  pi[1:2] ~ ddirch(gamma[])
  tau ~ dgamma(11, 280)
  lambda.vec[2] ~ dgamma(11, 70)

  theta.vec[2] <- 1/lambda.vec[2]
  diff <- 1/tau
  theta.vec[1] <- theta.vec[2] + diff
  lambda.vec[1] <- 1/theta.vec[1]
```

```
19
20
21 }
```

B4_1.R

B.2 Proposed response adaptive randomization procedure for survival response

```
1  survival.randomization <- function(isim.start, # start from zero
2                                     seed, # used in set.seed function
3                                     nSims, # number of simulations
4                                     nSbjs, # total number of patients in the trial
5                                     n.init, # number of subjects in the equal
                                          randomizaition
6                                     alpha, # nonnegative number in the DBCD procedure
7                                     weight1.A, weight2.A,
8                                     weight1.B, weight2.B,
9                                     theta1.A, theta2.A, # mean survival time
10                                    theta1.B, theta2.B,
11                                    R, # patient arrival times follow uniform dist on
                                           [0, R]
12                                    D,
13                                    n.iter, n.thin, n.update) {
14
15   allocation.prob.matrix = matrix(, nrow=nSims-isim.start, ncol=nSbjs)
16
17   p.val.vec = {}
18   z.score.vec = {}
19   num.event.vec = {}
20   total.obs.surv.time.vec = {}
21   allocation.prop.final.vec = {}
```

```
N.A.final.vec = {}
N.B.final.vec = {}
convergence.A.vec = {}
convergence.B.vec = {}

convergence.theta1.A.vec = {}
convergence.theta2.A.vec = {}
convergence.pi1.A.vec = {}
convergence.pi2.A.vec = {}

convergence.theta1.B.vec = {}
convergence.theta2.B.vec = {}
convergence.pi1.B.vec = {}
convergence.pi2.B.vec = {}

epsilon.final.A.vec = {}
epsilon.final.B.vec = {}

####### target allocation proportion function##############################
#rho <- function(Psi.A, Psi.B, variance.A, variance.B) {                  #
# sd.A = sqrt(variance.A)                                                 #
# sd.B = sqrt(variance.B)                                                 #
# rho = (sd.A*sqrt(Psi.B))/(sd.A*sqrt(Psi.B)+sd.B*sqrt(Psi.A))            #
# return(rho)                                                             #
#}                                                                         #
###########################################################################

######## the DBCD procedure
DBCD <- function(Psi.A, Psi.B, variance.A, variance.B, N.A, N.B) {
  if (is.na(Psi.A) | is.na(Psi.B)) {
    g.function = 0.5
  } else {
    sd.A = sqrt(variance.A)
    sd.B = sqrt(variance.B)
    x = N.A/(N.A+N.B)
```

```
      y = (sd.A*sqrt(Psi.B))/(sd.A*sqrt(Psi.B)+sd.B*sqrt(Psi.A))
      if (x==0) {
        g.function = 1
      } else if (x==1) {
        g.function = 0
      } else{
        g.function = y*(y/x)^alpha/(y*(y/x)^alpha+(1-y)*((1-y)/(1-x))^alpha)
      }
    }

      return(g.function)
}

######## expected proportion of event function
epsilon <- function(theta, D, R) {
  epsilon = 1 - theta/D + exp(-D/theta)*theta/(D*R)*(exp(R/theta)*(2*theta-R)-2*
      theta)
  return(epsilon)
}

jags.params <- c('lambda.vec', 'pi', 'theta.vec', 'tau')

set.seed(seed+isim.start)

for (isim in (isim.start+1):nSims) {

  surrogate = {}          # surrogate endpoint
  event.time = {}         # event time
  trt = {}
  #trt.A = {}
  #trt.B = {}
  censor.time = {}

  convergence.A = {}
  convergence.B = {}
```

```
91
92      convergence.theta1.A = {}
93        convergence.theta2.A = {}
94        convergence.pi1.A = {}
95      convergence.pi2.A = {}
96
97      convergence.theta1.B = {}
98        convergence.theta2.B = {}
99        convergence.pi1.B = {}
100     convergence.pi2.B = {}
101
102
103     #### first, generate the patient arrival time for all the patients
104     arrival.time.disorder <- runif(nSbjs, 0, R)
105     arrival.time = arrival.time.disorder[order(arrival.time.disorder)]
106
107     for (iSbj in 1:nSbjs) {
108
109       censor.time.interim = ifelse(censor.time < arrival.time[iSbj], censor.time,
              arrival.time[iSbj])
110       t.to.event.interim = pmin(event.time, censor.time.interim)
111       isCensored.interim = event.time > censor.time.interim
112       is.censored.interim = as.numeric(isCensored.interim)
113       is.event.interim = 1 - is.censored.interim
114       t.cen.interim <- t.to.event.interim
115       t.to.event.na.interim = t.to.event.interim
116       t.to.event.na.interim[is.censored.interim==1] <- NA
117
118       surrogate.interim.A = surrogate[trt==1]
119       surrogate.interim.B = surrogate[trt==2]
120
121       t.to.event.na.interim.A = t.to.event.na.interim[trt==1]
122       t.to.event.na.interim.B = t.to.event.na.interim[trt==2]
123
124       t.cen.interim.A = t.cen.interim[trt==1]
```

```
125       t.cen.interim.B = t.cen.interim[trt==2]
126
127       is.censored.interim.A = is.censored.interim[trt==1]
128       is.censored.interim.B = is.censored.interim[trt==2]
129
130       is.event.interim.A = is.event.interim[trt==1]
131       is.event.interim.B = is.event.interim[trt==2]
132
133          N.A.interim = length(surrogate.interim.A)
134          N.B.interim = length(surrogate.interim.B)
135
136          U = runif(1, 0, 1)
137
138          if (N.A.interim==0 | N.B.interim==0 | sum(is.event.interim.A) < 3 | sum(is.
                 event.interim.B) < 3) {
139           # equal randomization
140           allocation.prob.matrix[isim-isim.start, iSbj] = 0.5
141           if (U < 0.5) {
142             surrogate.A <- rbinom(1, 1, weight1.A)
143           surrogate.A = ifelse(surrogate.A==1, 1, 2)
144           surrogate = rbind(surrogate, surrogate.A)
145           event.time.A <- ifelse(surrogate.A==1, rexp(1, 1/theta1.A), rexp(1, 1/
                  theta2.A))
146           event.time = rbind(event.time, event.time.A)
147           trt = rbind(trt, 1)
148           } else {
149             surrogate.B <- rbinom(1, 1, weight1.B)
150           surrogate.B = ifelse(surrogate.B==1, 1, 2)
151           surrogate = rbind(surrogate, surrogate.B)
152           event.time.B <- ifelse(surrogate.B==1, rexp(1, 1/theta1.B), rexp(1, 1/
                  theta2.B))
153           event.time = rbind(event.time, event.time.B)
154           trt = rbind(trt, 2)
155           }
156          } else {
```

```
157           # adaptive randomization
158           ####### initial values for the Bayesian model
159         yInit.A = rep(NA, length(surrogate.interim.A))
160         yInit.A[is.censored.interim.A==1] = t.cen.interim.A[is.censored.interim.A
                ==1]+1
161
162         inits.A <- list(list(pi=c(0.7, 0.3), lambda.vec=c(NA, 0.02), tau=0.03, t.to.
                event=yInit.A, .RNG.seed=12345, .RNG.name="base::Mersenne-Twister"),
163                         list(pi=c(0.8, 0.2), lambda.vec=c(NA, 0.03), tau=0.02, t.to.
                            event=yInit.A, .RNG.seed=1234, .RNG.name="base::Mersenne-
                            Twister"),
164                         list(pi=c(0.5, 0.5), lambda.vec=c(NA, 0.05), tau=0.01, t.to.
                            event=yInit.A, .RNG.seed=123456, .RNG.name="base::Mersenne
                            -Twister"))
165
166         gibbs.data.A <- list(t.to.event = t.to.event.na.interim.A, t.cen = t.cen.
                interim.A, is.censored = is.censored.interim.A, N = length(t.to.event.na
                .interim.A), surrogate = surrogate.interim.A, gamma = c(0.5, 0.5))
167
168         gibbs.jags.A <- jags(data = gibbs.data.A, inits = inits.A, parameters.to.
                save = jags.params, n.iter = n.iter, model.file = 'survival.bug', n.thin
                =n.thin)
169
170         jagsfit.upd.A <- autojags(gibbs.jags.A, n.update=n.update)
171         myfit.A <- as.mcmc(jagsfit.upd.A)
172         out.A <- summary(myfit.A)
173
174         myfit.list.A <- mcmc.list(myfit.A)
175         gelman.A <- gelman.diag(myfit.list.A, multivariate=FALSE)
176         convergence.theta1.A.interim <- ifelse(gelman.A$psrf[7,1]<=1.1, 0, 1)
177         convergence.theta2.A.interim <- ifelse(gelman.A$psrf[8,1]<=1.1, 0, 1)
178         convergence.pi1.A.interim <- ifelse(gelman.A$psrf[4,1]<=1.1, 0, 1)
179         convergence.pi2.A.interim <- ifelse(gelman.A$psrf[5,1]<=1.1, 0, 1)
180         convergence.A.psrf <- ifelse(gelman.A$psrf[c(2,3,6),1]<=1.1, 0, 1)
181         convergence.A.ind <- sum(convergence.A.psrf)
```

```
182         convergence.A <- rbind(convergence.A, convergence.A.ind)
183         convergence.theta1.A <- rbind(convergence.theta1.A, convergence.theta1.A.
                interim)
184         convergence.theta2.A <- rbind(convergence.theta2.A, convergence.theta2.A.
                interim)
185         convergence.pi1.A <- rbind(convergence.pi1.A, convergence.pi1.A.interim)
186         convergence.pi2.A <- rbind(convergence.pi2.A, convergence.pi2.A.interim)
187
188         theta.1.interim.A = out.A$statistics[7,1]
189         theta.2.interim.A = out.A$statistics[8,1]
190         weight.1.interim.A = out.A$statistics[4,1]
191         weight.2.interim.A = out.A$statistics[5,1]
192         theta.interim.A = weight.1.interim.A*theta.1.interim.A+weight.2.interim.A*
                theta.2.interim.A  ## mean survival time
193         variance.interim.A = theta.interim.A^2
194         epsilon.interim.A = epsilon(theta=theta.interim.A, D=D, R=R)
195         variance.interim.censor.A = variance.interim.A/epsilon.interim.A
196         N.A.interim = length(surrogate.interim.A)
197
198         ## for treatment B
199         ####### initial values for the Bayesian model
200         yInit.B = rep(NA, length(surrogate.interim.B))
201         yInit.B[is.censored.interim.B==1] = t.cen.interim.B[is.censored.interim.B
                ==1]+1
202
203         inits.B <- list(list(pi=c(0.7, 0.3), lambda.vec=c(NA, 0.02), tau=0.03, t.to.
                event=yInit.B, .RNG.seed=12345, .RNG.name="base::Mersenne-Twister"),
204                         list(pi=c(0.8, 0.2), lambda.vec=c(NA, 0.03), tau=0.02, t.to.
                                event=yInit.B, .RNG.seed=1234, .RNG.name="base::Mersenne-
                                Twister"),
205                         list(pi=c(0.5, 0.5), lambda.vec=c(NA, 0.05), tau=0.01, t.to.
                                event=yInit.B, .RNG.seed=123456, .RNG.name="base::Mersenne
                                -Twister"))
206
```

```
207         gibbs.data.B <- list(t.to.event = t.to.event.na.interim.B, t.cen = t.cen.
                interim.B, is.censored = is.censored.interim.B, N = length(t.to.event.na
                .interim.B), surrogate = surrogate.interim.B, gamma = c(0.5, 0.5))
208
209         gibbs.jags.B <- jags(data = gibbs.data.B, inits = inits.B, parameters.to.
                save = jags.params, n.iter = n.iter, model.file = 'survival.bug', n.thin
                =n.thin)
210
211         jagsfit.upd.B <- autojags(gibbs.jags.B, n.update=n.update)
212         myfit.B <- as.mcmc(jagsfit.upd.B)
213         out.B <- summary(myfit.B)
214
215         myfit.list.B <- mcmc.list(myfit.B)
216         gelman.B <- gelman.diag(myfit.list.B, multivariate=FALSE)
217         convergence.theta1.B.interim <- ifelse(gelman.B$psrf[7,1]<=1.1, 0, 1)
218         convergence.theta2.B.interim <- ifelse(gelman.B$psrf[8,1]<=1.1, 0, 1)
219         convergence.pi1.B.interim <- ifelse(gelman.B$psrf[4,1]<=1.1, 0, 1)
220         convergence.pi2.B.interim <- ifelse(gelman.B$psrf[5,1]<=1.1, 0, 1)
221         convergence.B.psrf <- ifelse(gelman.B$psrf[c(2,3,6),1]<=1.1, 0, 1)
222         convergence.B.ind <- sum(convergence.B.psrf)
223         convergence.B <- rbind(convergence.B, convergence.B.ind)
224         convergence.theta1.B <- rbind(convergence.theta1.B, convergence.theta1.B.
                interim)
225         convergence.theta2.B <- rbind(convergence.theta2.B, convergence.theta2.B.
                interim)
226         convergence.pi1.B <- rbind(convergence.pi1.B, convergence.pi1.B.interim)
227         convergence.pi2.B <- rbind(convergence.pi2.B, convergence.pi2.B.interim)
228
229         theta.1.interim.B = out.B$statistics[7,1]  # mean survival time
230         theta.2.interim.B = out.B$statistics[8,1]
231         weight.1.interim.B = out.B$statistics[4,1]
232         weight.2.interim.B = out.B$statistics[5,1]
233         theta.interim.B = weight.1.interim.B*theta.1.interim.B + weight.2.interim.B*
                theta.2.interim.B  ## mean survival time
234         variance.interim.B = theta.interim.B^2
```

```
235         epsilon.interim.B = epsilon(theta=theta.interim.B, D=D, R=R)
236         variance.interim.censor.B = variance.interim.B/epsilon.interim.B
237         N.B.interim = length(surrogate.interim.B)
238
239         ## update the allocation ratio before randomize the next subject
240         Psi.interim.A = 1/theta.interim.A
241         Psi.interim.B = 1/theta.interim.B
242         allocation.interim = DBCD(Psi.A=Psi.interim.A, Psi.B=Psi.interim.B, variance
                .A=variance.interim.censor.A, variance.B=variance.interim.censor.B, N.A=
                N.A.interim, N.B=N.B.interim)
243
244         allocation.prob.matrix[isim-isim.start, iSbj] = allocation.interim
245
246         if (U < allocation.interim) {
247           ## allocate this patient to trt A
248           surrogate.A <- rbinom(1, 1, weight1.A)
249           surrogate.A = ifelse(surrogate.A==1, 1, 2)
250           surrogate = rbind(surrogate, surrogate.A)
251           event.time.A <- ifelse(surrogate.A==1, rexp(1, 1/theta1.A), rexp(1, 1/
                theta2.A))
252           event.time = rbind(event.time, event.time.A)
253           trt = rbind(trt, 1)
254         } else {
255           ## allocate this patient to trt B
256           surrogate.B <- rbinom(1, 1, weight1.B)
257           surrogate.B = ifelse(surrogate.B==1, 1, 2)
258           surrogate = rbind(surrogate, surrogate.B)
259           event.time.B <- ifelse(surrogate.B==1, rexp(1, 1/theta1.B), rexp(1, 1/
                theta2.B))
260           event.time = rbind(event.time, event.time.B)
261           trt = rbind(trt, 2)
262         }
263          }
264          ## simulate the censoring time for this patient
265       censor.time.i = runif(1, min=0, max=D)
```

```
      censor.time = rbind(censor.time, censor.time.i)
    }

    ### update the data for the last subject
    censor.time.final = ifelse(censor.time < D, censor.time, D)
    t.to.event.final = pmin(event.time, censor.time.final)
    is.censored.final = ifelse(t.to.event.final==censor.time.final, 1, 0)
    is.event.final = 1 - is.censored.final

    ### log rank test
    test <- survdiff(Surv(t.to.event.final, is.event.final) ~ trt, rho = 0)
    p.val <- 1 - pchisq(test$chisq, length(test$n) - 1)
    p.val.vec <- rbind(p.val.vec, p.val)

    ### number of events in the trial
    num.event <- sum(is.event.final)
    num.event.vec <- rbind(num.event.vec, num.event)

    ### total observed survival time
    total.obs.surv.time <- sum(t.to.event.final*is.event.final)
    total.obs.surv.time.vec <- rbind(total.obs.surv.time.vec, total.obs.surv.time)

    ##############################################################################
    ########### point estimate of hazard based on primary endpint ###############
    ##############################################################################
    t.to.event.final.A <- as.vector(t.to.event.final)[trt==1]
    t.to.event.final.B <- as.vector(t.to.event.final)[trt==2]
    is.event.final.A <- as.vector(is.event.final)[trt==1]
    is.event.final.B <- as.vector(is.event.final)[trt==2]
    N.A.final <- length(t.to.event.final.A)
    N.B.final <- length(t.to.event.final.B)

    N.A.final.vec <- rbind(N.A.final.vec, N.A.final)
    N.B.final.vec <- rbind(N.B.final.vec, N.B.final)

```

```
301     ### maximum likelihood estimator
302     theta.marginal.final.A <- sum(t.to.event.final.A)/sum(is.event.final.A)
303     theta.marginal.final.B <- sum(t.to.event.final.B)/sum(is.event.final.B)
304
305     epsilon.final.A <- epsilon(theta=theta.marginal.final.A, D, R)
306     epsilon.final.B <- epsilon(theta=theta.marginal.final.B, D, R)
307
308     epsilon.final.A.vec <- rbind(epsilon.final.A.vec, epsilon.final.A)
309     epsilon.final.B.vec <- rbind(epsilon.final.B.vec, epsilon.final.B)
310
311     ### allocation proportion
312     allocation.prop.final <- N.A.final/(N.A.final+N.B.final)
313     allocation.prop.final.vec <- rbind(allocation.prop.final.vec, allocation.prop.
            final)
314
315     ### wald test
316     z.score <- (theta.marginal.final.A-theta.marginal.final.B)/sqrt(theta.marginal.
            final.A^2/sum(is.event.final.A)+theta.marginal.final.B^2/sum(is.event.final.
            B))
317     z.score.vec <- rbind(z.score.vec, z.score)
318
319     ### convergence
320     convergence.A.vec <- rbind(convergence.A.vec, sum(convergence.A))
321     convergence.B.vec <- rbind(convergence.B.vec, sum(convergence.B))
322
323     convergence.theta1.A.vec <- rbind(convergence.theta1.A.vec, sum(convergence.
            theta1.A))
324     convergence.theta2.A.vec <- rbind(convergence.theta2.A.vec, sum(convergence.
            theta2.A))
325     convergence.pi1.A.vec <- rbind(convergence.pi1.A.vec, sum(convergence.pi1.A))
326     convergence.pi2.A.vec <- rbind(convergence.pi2.A.vec, sum(convergence.pi2.A))
327
328     convergence.theta1.B.vec <- rbind(convergence.theta1.B.vec, sum(convergence.
            theta1.B))
```

```
329     convergence.theta2.B.vec <- rbind(convergence.theta2.B.vec, sum(convergence.
            theta2.B))
330     convergence.pi1.B.vec <- rbind(convergence.pi1.B.vec, sum(convergence.pi1.B))
331     convergence.pi2.B.vec <- rbind(convergence.pi2.B.vec, sum(convergence.pi2.B))
332   }
333
334   test.result = ifelse(p.val.vec < 0.05, 1, 0)
335   power = sum(test.result)/length(test.result)
336   N.A.est = mean(N.A.final.vec)
337   N.B.est = mean(N.B.final.vec)
338   num.event.est = mean(num.event.vec)
339   total.obs.surv.time.est = mean(total.obs.surv.time.vec)
340   allocation.prop.est = mean(allocation.prop.final.vec)
341   epsilon.A.est = mean(epsilon.final.A.vec)
342   epsilon.B.est = mean(epsilon.final.B.vec)
343
344   list(power=power, N.A.est=N.A.est, N.B.est=N.B.est, num.event.est=num.event.est,
          total.obs.surv.time.est=total.obs.surv.time.est, allocation.prop.est=
          allocation.prop.est, epsilon.A.est=epsilon.A.est, epsilon.B.est=epsilon.B.est,
           p.val.vec=p.val.vec, N.A.final.vec=N.A.final.vec, N.B.final.vec=N.B.final.vec
          , allocation.prop.final.vec=allocation.prop.final.vec, convergence.A.vec=
          convergence.A.vec, convergence.B.vec=convergence.B.vec, num.event.vec=num.
          event.vec, total.obs.surv.time.vec=total.obs.surv.time.vec, epsilon.final.A.
          vec=epsilon.final.A.vec, epsilon.final.B.vec=epsilon.final.B.vec, convergence.
          theta1.A.vec=convergence.theta1.A.vec, convergence.theta2.A.vec=convergence.
          theta2.A.vec, convergence.pi1.A.vec=convergence.pi1.A.vec,
345   convergence.pi2.A.vec=convergence.pi2.A.vec, convergence.theta1.B.vec=convergence.
          theta1.B.vec, convergence.theta2.B.vec=convergence.theta2.B.vec, convergence.
          pi1.B.vec=convergence.pi1.B.vec,
346   convergence.pi2.B.vec=convergence.pi2.B.vec, allocation.prob.matrix=allocation.
          prob.matrix, z.score.vec = z.score.vec)
347
348
349 }
```

B4_2.R

Vita

Hui Wang was born on February 8, 1989 in Gaomi, Shandong province, People's Republic of China. She received her Bachelor of Science Degree in Statistics from Qingdao University in 2011. She came to the United States and began her graduate education in the Department of Biostatistics at Virginia Commonwealth University in Richmond, Virginia in the fall of 2011.