8-2020

COVARIATE-ADAPTIVE AND RESPONSE-ADAPTIVE CLINICAL TRIAL DESIGNS

LU WANG

COVARIATE-ADAPTIVE AND RESPONSE-ADAPTIVE CLINICAL TRIAL DESIGNS

by

LU WANG, MS

APPROVED:

HONGJIAN ZHU, PHD

DAVID R. LAIRSON, PHD

SHREELA V. SHARMA, PHD

JING NING, PHD

DEAN, THE UNIVERSITY OF TEXAS
SCHOOL OF PUBLIC HEALTH

DEDICATION

To my families

COVARIATE-ADAPTIVE AND RESPONSE-ADAPTIVE CLINICAL TRIAL DESIGNS

by

LU WANG
BS, University of Science and Technology of China, 2009
MS, Brown University, 2011

Presented to the Faculty of The University of Texas

School of Public Health

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS
SCHOOL OF PUBLIC HEALTH
Houston, Texas
August, 2020

ACKNOWLEDGEMENTS

I would like to thank the following people, without whom I would not have been able to complete this research. My advisor Dr. Hongjian Zhu, for his enthusiasm for research, for his support, encouragement, and patience during the last several years. Without his guidance and persistent help, this dissertation would not have been possible. His passion for research and life was contagious and motivational and will continue to benefit my future career and life.

I would like to thank my committee members: minor advisor Dr. David R. Lairson and breadth advisor Dr. Shreela V. Sharma. I am grateful for their helpful discussion and comments. I would like to thank Dr. Jing Ning for her advice and inspiration for my research. I would like to thank Dr. Barbara C. Tilley for providing me the great opportunity of working on the MDS-UPDRS and UDysRS studies.

To my friends, my parents, I am forever grateful for your love, patience and understanding.

Finally, to my husband, Mingqiang: your love and understanding helped me through the dark times. And my daughter, Olivia without whom this work would have been completed three years earlier.

COVARIATE-ADAPTIVE AND RESPONSE-ADAPTIVE CLINICAL TRIAL DESIGNS

Lu Wang, MS, PhD
The University of Texas
School of Public Health, 2020

Dissertation Supervisor: Hongjian Zhu, PhD

Most of the research on Response-adaptive randomization(RAR) designs has been focused on clinical trials with a single endpoint. However modern clinical trials are often complex, with multiple competing objectives and multiple endpoints. We overcome the obstacles introduced by the large number of unknown parameters and the possible correlations between the multiple endpoints. We obtained the optimal allocation proportions for the following two optimization problems: (1) maximizing the power of the test of homogeneity with a fixed sample size, and (2) minimizing the expected weighted number of failures with a fixed power. Further, we implemented these optimal allocations through response-adaptive randomization procedures. Our theoretical results provided the foundation for the implementation and further investigation of the procedure, and our numerical studies demonstrated its ability to achieve diverse objectives.

Covariate adaptive randomization (CAR) designs including the stratified permuted block randomization design is a standard in clinical trials. It is well accepted through numerous numerical studies that the type I error rate would be conservative if not all the randomization covariates were included in the data analysis following CAR designs. But the theoretical investigation for clinical trials using CAR designs for randomization and time-to-event outcomes for data analysis is lacking in the literature. In this paper, we proposed the test statistics, and demonstrated the effect of CAR designs on the type I error rate and power for such trials with simulations. We also proposed approaches

to control the type I error rate. Numerical studies demonstrated our showed that our proposed methods successfully protected the type I error rate in these trials. These numerical results offered practical guidance for future clinical trials employing CAR designs and survival analysis.

KEY WORDS: Clinical trial; Multiple endpoints; Optimal allocation; Power; Response-adaptive design; Accelerated failure time model; Covariate adaptive design; Conservative tests; Pocock and Simon's marginal procedure; Stratified permuted block design, Type I error.

Contents

List of Tables

List of Figures

Chapter 1

Backgrounds

Randomization is an essential part of clinical trials which ensures comparability of treatment groups, mitigates selection bias or other accidental bias in the design. Randomization produces comparable intervention groups to avoid confounding from other known or unknown factors and balances prognostic variables. It ensures each patient have an equal chance to receive any one of the treatment thus the groups are alike except for the treatments. It also reduces other forms of experimenter and subject biases by using blinding or masking. With randonmization probability theory can be used to express the likelihood that whether the difference in outcome between treatment groups are merely random.

In practice permuted block design is the most popular randomization procedure in which randomization is applied in blocks. After randomization some fixed allocation ratio typically 1:1 in two-arm trials is targeted. When there are multiple experimental objectives are pursued in the clinical trials this easy-to-implement procedure may not be optimal. Adaptive designs have long been proposed to deal with the ethical problem of randomly assigning a treatment to patients in clinical trials.

Many procedures have been proposed for to randomly assign participants to treatment groups in clinical trials. In this chapter, an overview of randomization procedures in

this dissertation is detailed. Non-adaptive designs are discussed in 1.1 which is the foundation of the latter discussed adaptive designs. Section 1.2 reviewed different adaptive randomization procedures . Chapter 2 and 3 provides the oprtimal allocations that maximize power for a fixed sample size or minimize expected number of failures expanding upon Response-adaptive randomization (RAR) designs discussed in Section 1.2. Chapter 4 provides extensive simulation studies in the impact in inference and hypothesis testing when Covariate-adjusted randomization is used.

1.1 Non-adaptive designs

Non-adaptive randomization procedures assign the participants to the treatment groups at random and do not consider how other subjects have responded so far in the trial or the participants' baseline information. These methods are commonly used when it is desired to balance the treatment assignments that is to have almost equal patients in each of the treatment group.

1.1.1 Complete Randomization

In clinical trials using complete randomization procedure the treatment assignments are usually identical and independently distributed from Bernoulli distribution with success probability 1/2. Since the treatments are generated compete randomly there is no selection bias.

However Rosenberger and Lachin (2015) showed with simulation that there is still a non-negligible probability of some imbalances between treatments as well as a small probablity of severe imbalances. Akthough the imbalance will still give us unbiased estimation of treatment group difference it will lead to less precision of the estimation hence decrease the power. To mitigate this disadvantage of complete randomization many restricted

randomization procedures are proposed.

1.2 Adaptive randomization procedures of clinical trials

1.2.1 Restricted Randomization

In restricted randomization, patients arrive sequentially and the probability of the future treatment assignment is chosen based on the past treatment assignments to achieve the balance of objects across treatment groups. Many methods have been proposed to pretect agiast biases.

1.2.2 Forced Balance Procedures

Random Allocation Rule

Random Allocation Rule randomly choose a subset of $n/2$ out of n and assign to group A, the remainder to group B. Thus, the sample sizes in each group, say N_A and N_B, each equal to $1/2$. It's an urn with $n/2$ A balls and $n/2$ B balls, N draws without replacement. Consider a clinical trial with two groups A and B. Let $F_n = \{I_1, ..., I_n\}$ be a set of treatment assignments for the first n stages of the randomization process. Then denote the probability that A is assigned to the jth patient, conditional on the first $j-1$ treatment assignments ϕ_j Here

$$\phi_j = E\left(I_j | F_{j-1}\right) = \frac{N/2 - N_A(j-1)}{n-(j-1)} \tag{1.1}$$

Lachin (1988) showed that random allocation rule yields a substantial potential for selection bias in an unmasked trial and there is a greater likelihood of a covariate imbalance

with the random allocation rule than with complete randomization.

Truncated Binomial Design (TBD)

The Truncated Binomial Design consists of tossing a fair coin for the allocation of patients until $n/2$ heads or tails have occurred. And then the remainder is filled with the opposite treatment with probability . Then truncated binomial design allocation rule is defined by:

$$\phi_j = E\left(I_j|F_{j-1}\right) = \begin{cases} 1/2, & \text{if } max\left\{N_A(j-1), N_B(j-1)\right\} < n/2, \\ 0, & \text{if } N_A(j-1) = n/2, \\ 1, & \text{if } N_B(j-1) = n/2. \end{cases} \tag{1.2}$$

Note that TBD results in sequences that achieve final balance but are not equiprobable. The covariates in the treatment groups can be imblance.

Permuted Block Design (PBD)

The permuted-blockdesign involves randomizing patients to treatment groups in sequential blocks. For example, in a design where the block size is four, there are six possible ways to make treatment assignments for a block: AABB, BBAA, ABAB, BABA, ABBA, and BAAB. Blocks of even size 2b are filled using either a random allocation rule or a truncated binomial design. The maximum imbalance at any time point is then half a block size, b. Let R_j be the position patient j takes within his block. If we fill blocks using Random Allocation Rule, the allocation rule is:

$$\phi_j = E\left(I_j|F_{j-1}, b, R_j\right) = \frac{b - \sum_{l=j+1-R_j}^{j-1} I_l}{2b - R_j + 1} \tag{1.3}$$

This design periodically achieves balance between the number of patients in each treatment group. And there is no treatment imbalance if the total sample size is a multiple of the block size. However under this setting every block has at least one deterministic

assignment. Matts and Lachin (1988) also showed that PBD has an increased probability of selection bias and accidental bias due to the achievementof periodic balance.

Random Block Design

In Random Block Design, blocks of size $2, 4, 6, ..., 2K$ are randomly selected and equiprobable. The randomness will decrease the predictability of the treatment assignmen since the block size not a constant and no assignments are made with probability 1. The maximum imbalance is K, there is terminal balance if the last block is filled. Let B_j be the block size of the block with the jth patient, Let R_j define the position patient j takes within his block, ranging from $1, ..., B_j$. Each block can be filled with any forced balance procedure. If we fill blocks using Random Allocation Rule, the allocation rule is:

$$\phi_j = E\left(I_j | F_{j-1}, b, R_j\right) = \frac{B_j/2 - \sum_{l=j+1-R_j}^{j-1} I_l}{B_j - R_j + 1} \tag{1.4}$$

Matts and Lachin (1988) also showed that Random Block Design can reduce the potential for selection bias. But the only way to completely eliminate it is to randomize patients as a block rather than individually as they arrive for entry into the trial.

Biased Coin Design (BCD) Biased Coin Design was first proposed by Efron (1971) to ensure a balanced experiment while still retain some randomness. It minimizes the possibility of selection bias and accidental bias. Many extensions have been developed since then.

Efron's Biased-Coin Design

In Efron's biased-coin design the allocations depend on N_A and N_B through the difference D_n. It gives a higher probability $p > 1/2$ of assigning the treatment that has the fewest

assignments thus far. Then the allocation rule is defined by:

$$\phi_j = E\left(I_j|F_{j-1}\right) = \begin{cases} 1/2, & \text{if } D_{j-1} = 0, \\ p, & \text{if } D_{j-1} < 0, \\ 1-p, & \text{if } D_{j-1} > 0. \end{cases} \tag{1.5}$$

where $0.5 < p <= 1$. Whn $p = 0.5$ it becomes complete randomization with no control over the balance. The parameter p represents a trade-off between balance and predictability. Efron recommended $p = 2/3$, which can easily be implemented using a six sided die.

Big Stick Design (BSD)

BSD is potential alternatives to the PBD (Zhao (2014)). It has no block issue and is only resticted by the two boundaries formed by a pre-specified maximal tolerated imbalance (MTI). Let δ =MTI, the allocation rule is defined by:

$$\phi_j = E\left(I_j|F_{j-1}\right) = \begin{cases} 0, & \text{if } D_{j-1} = \delta, \\ 0.5, & \text{if } |D_{j-1}| < \delta, \\ 1, & \text{if } D_{j-1} = -\delta. \end{cases} \tag{1.6}$$

Block Urn Design (BUD)

Zhao and Weng (2011) proposed BUD as alternatives to the PBD using an active urn and an inactive urn. Let $b = \lambda \boldsymbol{W}$ be the block size, and λ be the number of minimal balanced sets in each block, $\boldsymbol{W} = \sum w_j$ with w_j assignments for treatment j. The allocation rule is defined by:

$$\phi_j = E\left(I_j|F_{j-1}\right) = \frac{1}{2} - \frac{D_{j-1}}{4\lambda - 2\left|D_{j-i}\right|} \tag{1.7}$$

Wei's Urn Design

Wei et al. (1979) proposed a generalized Polya's urn design for sequential medical trials.

Consider an urn contains α white and α red balls originally. A ball is drawn at random and replaced for a treatment assignment. Treatment A is assigned if the ball is white otherwise treatment B is assigned. ,add β additional balls of the opposite color of the ball chosento the urn. This urn design is designated by UD(α, β). The allocation rule is

$$\phi_j = E\left(I_j|F_{j-1}\right) = \frac{\alpha + \beta N_A(j-1)}{2\alpha + \beta(j-1)} \tag{1.8}$$

Wei and Lachin (1988) showed that the urn design can reduce experimentalbias better than other restricted randomization procedures.

Generalized Biased Coin Design (GBCD)

Smith (1984) introduced a more general class of allocation procedures given by

$$\phi_j = E\left(I_j|F_{j-1}\right) = \frac{N_B^{\rho}(j-1)}{N_A^{\rho}(j-1) + N_B^{\rho}(j-1)} \tag{1.9}$$

where ρ is a tuning parameter. Wheb $\rho = 0$ Smith's design is complete randomization, as ρ approaches to infinity, it is becomes the PBD with a block size $b = 2$, and if $\rho = 1$ it is Wei's urn design.

Accelerated Biased Coin Design (ABCD)

Antognini and Giovagnoli (2004) proposed the ABCD method which adapts according to the magnitude of the imbalance of the biased coin design using a tuning parameter a. the allocation rule is defined by:

$$\phi_j = E\left(I_j|F_{j-1}\right) = \begin{cases} 1/2, & \text{if } D_{j-1} = 0, \\ \frac{|D_{j-1}|^a}{|D_{j-1}|^a+1}, & \text{if } D_{j-1} \leq -1, \\ \frac{1}{|D_{j-1}|^a+1}, & \text{if } D_{j-1} \geq 1. \end{cases} \tag{1.10}$$

When the tuning parameter $a = 0$ it is complete randomization. As a approaches to infinity, the ABCD is equivalent to the Big Stick Design with $b = 2$.

1.2.3 Response-adaptive randomization

The basic idea is to skew allocation probability according to the previous treatment assignments and responses in order to meet certain objectives such as maximize power and minimize expected treatment failures. Ivanova and Rosenberger (2000) showed that unequal allocation can sometimes result in gains in power and Rosenberger and Stallard gave the optimal allocation to minimize number of expected treatment failures for two treatments clinical trials.

Hu and Rosenberger (2003) formalized the development of optimal response-adaptive randomization procedures in a "3-step" approach. At the first step, the objective is chosen and an optimal allocation is derived as a solution to some formal optimization problem. A general framework of the optimization problem was given by Jennison and Turnbull (1999) triggered a series of work on optimal response-adaptive randomization designs for trials with binary outcomes (Rosenberger et al. (2001); Ivanova and Rosenberger (2001); Rosenberger and Hu (2004)), normal outcomes Biswas and Mandal (2004); Zhang and Rosenberger (2006); Gwise et al. (2008); Biswas and Bhattacharya (2009), Biswas and Bhattacharya (2010)), and survival outcomes (Zhang and Rosenberger (2007)).

Neyman Allocation

For two-treatment trials, Melfi and Page (1998) discussed an optimal allocation called *Neyman allocation* to maximize power as follow:

Consider an experiment with 2 treatments and binary outcomes, where the success probabilities on treatment 1 and treatment 2 for each endpoints are given by p_1, p_2 respectively and the failure probabilities are given by q_1, q_2. Treatments are to be applied

to n subjects, and treatment k will be applied to n_k subjects, k=2. To compare the two treatments the hypothesis testing is:

$$H_0 : p_1 - p_2 = 0 \quad \text{vs} \quad H_A : p_1 - p_2 \neq 0$$

Wald test is used with a pre-specified Type I error rate using the test statistic

$$Z = \frac{\hat{p_1} - \hat{p_2}}{\sqrt{\frac{\hat{p_1}\hat{q_1}}{n_1} + \frac{\hat{p_2}\hat{q_2}}{n_2}}}$$

To find the optimal allocation $R = n_1/n_2$ which minimizes the variance of the difference in sample proportions The optimal allocation proportions to maximize power are denoted as ρ_1, ρ_2, where $\rho_1 = n_1/n, \rho_2 = n_2/n$. To minimize the variance

$$f(\rho_1) = \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2} = \frac{1}{n}\left(\frac{p_1 q_1}{\rho_1} + \frac{p_2 q_2}{1 - \rho_1}\right)$$

Let

$$f'(\rho_1) = \frac{1}{n}\left(-\frac{p_1 q_1}{\rho_1^2} + \frac{p_2 q_2}{(1 - \rho_1)^2}\right) = 0$$

We have the solution

$$\rho_1^*(p_1, p_2) = \frac{\sqrt{p_1 q_1}}{\sqrt{p_1 q_1} + \sqrt{p_2 q_2}},\ \rho_2^*(p_1, p_2) = 1 - \rho_1^*(p_1, p_2).$$

This is known as the Neyman allocation. One disadvantage of Neyman allocation is that it would be unethical when $p_1 > q_2$ since it will assign more patients to the inferior treatment.

RSIHR Allocation

Rosenberger et al. (2001) proposed this optimal allocation to minimize the expected

number of treatment failures for trials with binary response. The allocation is given by

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

In the continuous case, the optimization problem becomes

$$\begin{aligned} &\min_{n_A, n_B} \mu_1 n_1 + \mu_2 n_2 \\ &s.t.\ \frac{\sigma_1^2}{n1} + \frac{\sigma_1^2}{n1} \geq M. \end{aligned}$$

Then the optimal allocation is given by

$$\rho_1^*(\mu_1, \mu_2, \sigma_1, \sigma_2) = \frac{\sqrt{\mu_1}\sigma_1}{\sqrt{\mu_1}\sigma_1 + \sqrt{\mu_2}\sigma_2},\ \rho_2^*(\mu_1, \mu_2, \sigma_1, \sigma_2) = 1 - \rho_1^*(\mu_1, \mu_2, \sigma_1, \sigma_2)$$

However, when $\mu_1 < \mu_2$ it is possible that $n_1/n_2 < 1/2$, that is RSIHR allocation may assign more patients to the inferior treatment.

Urn Allocation

Rosenberger (2002) showed many response-adaptive randomization procedures have been based on probabilistic properties of urn models most of which aiming to assign more patients to the better treatment. These include the randomized -the-winner (RPW) (Wei (1978)) and drop-the-loser (DL) rule (Ivanova (2003)). The allocation is given by

$$\rho_1^*(p_1, p_2) = \frac{q_1}{q_1 + q_2} = \frac{1 - p_1}{2 - p_1 - p_2},\ \rho_2^*(p_1, p_2) = 1 - \rho_1^*(p_1, p_2).$$

Biswas and Mandal Allocation

Biswas and Mandal (2004) considered the following optimization problem:

$$\min_{n_A,n_B} \left\{ n_1 \Phi\left(\frac{\mu_1 - c}{\sigma_1}\right) + n_1 \Phi\left(\frac{\mu_2 - c}{\sigma_2}\right) \right\}$$
$$s.t.\ \frac{\sigma_1^2}{n1} + \frac{\sigma_1^2}{n1} \geq M.$$

which minimize the total number of responses larger than a threshold c. The allocation is given by:

$$\rho_1^*(\mu_1, \mu_2, \sigma_1, \sigma_2) = \frac{\sigma_1 \sqrt{\Phi\left(\frac{\mu_2 - c}{\sigma_2}\right)}}{\sigma_1 \sqrt{\Phi\left(\frac{\mu_2 - c}{\sigma_2}\right)} + \sigma_2 \sqrt{\Phi\left(\frac{\mu_1 - c}{\sigma_1}\right)}},\ \rho_2^* = 1 - \rho_1^*$$

This can be interpreted as minimization of the total expected failures.

Bandyopadhyay and Biswas Allocation

Bandyopadhyay and Biswas (2001) considered a trial with continuous responses in which patients are heterogeneous with respect to some prognostic factors.When the observation of ith patient follows

$$Y_i = I_i \mu_1 + (1 - I_i)\mu_2 + x_i' \beta + \epsilon_i \ \ (i = 1, 2..., n)\,,$$

where the ϵ's are i.i.d $N(0, \sigma^2)$ random variables. The allocation is given by:

$$\rho_1^*(\mu_1, \mu_2) = \Phi\left(\frac{\mu_1 - \mu_2}{T}\right),\ \rho_2^*(\mu_1, \mu_2) = 1 - \rho_1^*(\mu_1, \mu_2).$$

where T is the tuning parameter. The paper suggested to start with a larger value of T and switch over to progressively smaller values at suitable stages.

1.2.4 Response-Adaptive designs

Many response adaptive designs have been proposed fro clinical trials. One major purpose is to target certain treatment allocation and at the same time to maintain an adequate level of randomness in the allocation process.

Sequential Maximum Likelihood Estimation Procedure (SMLE)

It updates the optimal allocation by substituting estimators from the current data for the unknown parameters. Consider a clinical trial with binary response, after $j-1$ patients have responded, we can get the estimation $\hat{p_1}(j-1), \hat{q_1}(j-1), \hat{p_2}(j-1), \hat{q_1}(j-1)$ and $\hat{\rho} = \rho(\hat{p_1}(j-1), \hat{p_2}(j-1))$, then randomly assign the next patient with probabilty $\hat{\rho}$. Under the conditions given by Melfi et al. (2001) it will converge to the optimal allocation ρ almost surely in n.

Doubly Adaptive Biased Designs (DBCD)

Eisele (1994) and Eisele and Woodroofe (1995) introduced the doubly adaptive biased coin designs (DBCD) that adapts allocations based on treatment group frequencies as well as on the outcome information. It can be thougnt as a generalization of SMLE procedure. Suppose we have a function g defined on $[0,1] \times [0,1]$ such that:

(a) g is jointly continuous,

(b) $g(r,r) = r$,

(c) $g(p,r)$ is strictly decreasing in p and strictly increasing in r on $(0,1)^2$,

(d) g has bounded derivatives in both arguments.

For the jth patient, g represents the closeness of $N_1(j-1)/(j-1)$.Hu and Zhang (2004) proposed a family of allocations with a positive interger γ:

$$g(x,\rho) = \frac{\rho(\rho/x)^\gamma}{\rho(\rho/x)^\gamma + (1-\rho)((1-\rho)/(1-x))^\gamma}$$
$$g(0,\rho) = 1,$$
$$g(1,\rho) = 0.$$

when $\gamma = 0, g(x,\rho) = \rho$ it reduces to SMLE. When $\gamma = \infty$ the variance of the procedure is minimized but the assignments are completely predictable. γ is the tuning parameter controling the trade off between variantion and degree of randomization. We will discuss about how to implement DBCD step by step in Chapter 2. Rosenberger and Hu (2004) showed that DBCD are the preferred procedures, and simulations show that these procedures yield a modest reduction in expected treatment failures while preserving power over complete randomization. (2001).

Efficient Randomized-Adaptive Design (ERADE)

Hu et al. (2009) proposed a new family of effient randomized-adaptive designs which is easy to apply for clinical trials with both iscrete and continuous responses. It starts with assigning first n_0 (usually $n_0 = 2$) subjects with restricted randomization. When $j(j \geq 2n_0)$ subjects' responses are observed. Let $\hat{\rho}$ be the estimator based on the j observations. Then the ERADE assigns the $(j+1)$th patient to treatment 1 with probability

$$p_{j+1} = \begin{cases} \alpha\hat{\rho}_j & \text{if } N_1(j)/j > \hat{\rho}_j, \\ \hat{\rho}_j & \text{if } N_1(j)/j = \hat{\rho}_j, \\ 1-\alpha(1-\hat{\rho}_j) & \text{if } N_1(j)/j < \hat{\rho}_j. \end{cases}$$

where $0 \leq \alpha < q$ is a tuning parameter that control the degree of randomization. Hu

et al. (2009) recommended to choose α between 0.4 and 0.7. With $\rho = 1/2$) as target allocation and $\alpha = 2/3$, ERADE is Efron's biased coin design.

1.2.5 Covariate-adaptive randomization

These designs can be used to balance the patients prognostic factors such as gender, age and disease status in each treatment arm. This will avoid the inaccuracy introduced into the estimation of treatment effect due to the poor balance in patients' characteristics(FDA, 2019). Many Covariate-adaptive designs were proposed for clinical trials other than complete randomization. These designs can be used to balance the patients prognostic factors such as gender, age and disease status in each treatment arm. This will avoid the inaccuracy introduced into the estimation of treatment effect due to the poor balance in patients' characteristics (FDA, 2019). Balance in prognostic factors across treatment is also desirable for trials with small sample size or trials that require subgroup analysis etc. (Toorawa et al. (2009)).

There are many randomization procedures available in the literature. One of the most commonly used methods is stratified randomization which determines the strata first with the covariates' levels and then do the permuted block randomization within each stratum. To ensure balance over a large number of covariates, the minimization procedure was proposed by Taves (1974) and Pocock and Simon (1975) minimizing a weighted average of marginal imbalances. Many other methods such as Atkinson's class of procedures based on optimal designs for homoscedastic linear models (Atkinson (1982)) were also proposed recently. An overview of covariate-adaptive randomization methods can be found in Rosenberger et al. (2008). The limitation of these classical designs were discussed in Hu et al. (2012) and they also proposed a generalized family of covariate-adaptive designs with their theoretical properties.

For a long time very little theoretical work was done about the inference and tests

associated with covariate-adaptive designs. Practitioners typically use the same statistical methods as the complete randomization designs which may lead to the change of Type I error of the test. Many work based on simulation results were done without theoretical proof. (see, e.g. Birkett (1985) ; Forsythe (1987); Aickin (2009);Weir and Lees (2003); Hagino et al. (2004)). Shao et al. (2010) gave the proof of the properties of two sample t-test of treatment effects based on covariate-adaptive biased coin design. Ma et al. (2015) discussed the inference of covariates under covariate-adaptive designs. All of these studies were focused on linear models while in practice a large number of trials with censored survival data.

1.2.6 Covariate-adjusted response adaptive randomization

A combination of Covariate-adaptive randomization and Response-adaptive randomization which incorporates sequentially history information of response and covariate data as well as the observed covariate information of the incoming patient into the assignment of treatment. In clinical trials that treatment failures are costly in terms of ethical costs with known covariates that are believed to be important in prognosis, such design is helpful.

1.3 Clinical trials with multiple endpoints

Comparison of two or more samples with multiple endpoints is a common statistical problem in biomedical research. The clinical objective is to show at least one positive effect in treatment group. The problem is complicated by the fact that endpoints are usually correlated.

Most attention has been focused on clinical trials with only one endpoint. A general solution to binary response experiments with K treatments were given by Tymofyeyev

et al. (2007). For continuous responses Zhu and Hu (2009) gave the optimal allocations for clinical trials with exponential responses. Comparison of two or more samples with multiple endpoints is a common statistical problem in biomedical research. The clinical objective is to show at least one positive effect in treatment group. The problem is complicated by the fact that endpoints are usually correlated. In a clinical trial of an analgesic drug used to relieve arthritic pain (Jennison and Turnbull (1993), 1993), the primary endpoints were a measure of the amount of pain relief experienced by the patient and possible effect on the arthritic condition of the joint. Since the drug's success in relieving pain may lead the patient to be less careful in protecting the joint, the two outcome measures might be related. Many methods were proposed to make multiple endpoint adjustment in clinical trials such as the Bonferroni procedure, where one can reject all individual hypotheses $H_i(i = 1, 2, ...n)$ with $P_i \leq \alpha/n$. However this approach becomes more conservative in clinical trials when endpoints are correlated. And clinical trials often have to be large In order to assure adequate power resulting in more patients to placebo treatment or ineffective treatment. The two-sample Hotelling's T-Squared which is the multivariate extension of the common two-group Student's t-test is also the classical method to handle multiple endpoints when comparing two groups or treatments. O'Brien (1984) also proposed OLS and GLS tests for multiple endpoints. Various extension of O'Brien's work have also been proposed (Pocock et al. (1987),Tang et al. (1989b),Tang et al. (1989a),Tsai and Kozial (1994),Tang et al. (1993),Lefkopoulou and Ryan (1993),Follmann (1995),Follmann (1996),Lauter (1996),Läuter et al. (1998)).

1.4 Clinical trials with survival outcomes and AFT model

Time-to-event data arise when the time elapsing before an event occurs is of interest. In statistics they are known as survival data, since death is often the event of interest, particularly in clinical trials on cancer and heart disease. One key distinction between survival times and other continuous data is that some of the events of interest may not occurred by the time the study ends for all patients. For these patients we do not know when or whether the patient will experience the event. These times are called censored times. Non-parametric model such as Kaplan-Meier estimator, semi-Parametric models such as Cox proportional hazards model, and parametric models such as accelerated failure time model (AFT) are commonly used in survival analysis.

The accelerated failure time model has been used for traditional survival data under independent right censoring as a useful alternative to the commonly used Cox's proportional hazards model when the underlying assumptions are violated (Ponnuraja and Venkatesan (2010)). In recent years inference procedures and their asymptotic properties have been intensively studied (Buckley and James (1979); Koul et al. (1981); Tsiatis (1990); Miller and Halpern (1982); Ritov (1990); Lai and Ying (1991); Ying (1993); Lin and Ying (1995); Jin et al. (2003), 2003; Cai et al., 2009). In the clinical trial conducted by Brem et al. (1995), the effectiveness of biodegradable polymers impregnated with carmustine to treat recurrent malignant gliomas was evaluated. Proportional hazards regression model was used to adjust the survival curves. How ever the assumption of constant proportionality may not be satisfied as time progresses. Several approaches can be employed to estimate the parameters in the AFT model including Buckley–James es-

timator, rank-based estimator and the weighted least squares approach estimator (Stute (1993), Stute (1996)), which is equivalent to the inverse probability weighting (IPW) estimator.

1.5 Research Questions and Specific Aims

1.5.1 Specific Aims for Paper 1

We obtain the optimal allocation proportions for maximizing the power of the test of homogeneity with a fixed sample size. Further, we implement this optimal allocations through response-adaptive randomization procedures. Traditional framework for the optimization problem was applied to experimental designs for trials with a single endpoint. Testing of multiple endpoints will introduce large number of unknown parameters and the possible correlations between the multiple endpoints. Our theoretical results provide the foundation for the implementation and further investigation of the procedure, and our numerical studies demonstrate its ability to achieve the objective.

1.5.2 Specific Aims for Paper 2

We obtain the optimal allocation proportions for minimizing the expected weighted number of failures with a fixed power. Further, we implement this optimal allocations through response-adaptive randomization procedures. Simulation studies were performed to show that the proposed allocation performs well. We apply the method to a trial of perioperative total parenteral nutrition (TPN) treatment using the RAR design to demonstrate improvement of our proposed method from traditional design.

1.5.3 Specific Aims for Paper 3

We investigate the statistical inference, especially the control of type I error rate, of a clinical trial with AFT model and CAR designs. Investigation for clinical trials using CAR designs for randomization and time-to-event outcomes for data analysis using AFT models is lacking in the literature. We also proposed approaches to control the type I error rate. Numerical results offered practical guidance for future clinical trials employing CAR designs and survival analysis.

Chapter 2

Implementing Optimal Allocation in Clinical Trials with Multiple Endpoints to maximize the power of the test of homogeneity with a fixed sample size

2.1 Introduction

Clinical trials are complex experiments involving human beings. They often have multiple competing objectives, such as minimizing the total number of failures and maximizing the power of detecting treatment effects. These cannot be satisfied by traditional clinical trial designs that aim to balance the patient numbers across different treatments. Response-adaptive randomization (RAR) procedures achieve these objectives by skewing the allocation probability according to the previous treatment assignments and responses. Most of the research on RAR designs has been focused on clinical trials with a single endpoint. However, in modern clinical trials multiple possibly correlated endpoints are often evaluated simultaneously, and the clinical goal is to show a positive effect for at least one endpoint. There are different rationales for multiple (composite) endpoints. First, several different endpoints may be important for the participants. Second, the investigators are not always sure which outcome indicates a treatment effect (Friedman

et al. (2010)). Third, embedding multiple endpoints into the analysis can help to distinguish weaker signals of treatment effects from the noises of sampling errors (Moyé (2003)). Fourth, the total event rate can be increased, which can lead to a reduction in the required sample size.

Numerous real trials currently use multiple endpoints. For example, pulmonary function, neuro-psychological status, quality of life, and mortality were all assessed in the Nocturnal Oxygen Therapy Trial (Nocturnal Oxygen Therapy Trial Group (1980)); the trial compared continuous oxygen (O_2) therapy and 12-hour nocturnal O_2 therapy for chronic obstructive pulmonary disease. The Urokinase Pulmonary Embolism Trial (Urokinase Pulmonary Embolism Trial Study Group (1974)) explored the effect of three treatments, 12 hours of urokinase, 24 hours of urokinase, and 24 hours of streptokinase, on pulmonary embolism. They considered three endpoints: angiographic severity, lung scan perfusion defects, and hemodynamic variables. Roberts et al. (1984) investigated the effect of propranolol in limiting the myocardial infarct size. Treatment (Propranolol) or placebo was given intravenously upon randomization followed by orally for nine days to keep the heart rate between 45 and 60 beats per minute. The endpoints were infarct size as estimated from plasma MB creatine kinase activity, extent of area involved in pyrophosphate uptake, R-wave loss on electrocardiograms. In a clinical trial of an analgesic drug used to relieve arthritic pain (Jennison and Turnbull (1993)), the primary endpoints were a measure of the pain relief experienced by the patient and the effect on the arthritic condition of the joint. Since the drug's success in relieving pain may lead the patient to be less careful of the joint, the two outcome measures might be related. These examples motivate our investigation of the use of RAR designs to find optimal allocations for clinical trials with multiple endpoints; the goal is to achieve efficient objectives.

The idea underlying RAR designs can be traced back to Thompson (1933) and Robbins (1952). The play-the-winner rule (Zelen (1969)) and the randomized play-the-winner

rule (Wei and Durham (1978)) can reduce the number of patients receiving inferior treatments. Rosenberger et al. (2001) proposed an optimal allocation that minimizes the total number of failures while fixing the power. Ivanova and Rosenberger (2000) showed that an unequal allocation can sometimes result in a gain in the power. Ivanova (2003) proposed the drop-the-loser rule that achieves minimal variability. Hu and Rosenberger (2003) proved that RAR designs can increase statistical efficiency in certain clinical trials. A comprehensive introduction to RAR designs can be found in Hu and Rosenberger (2006).

Hu and Rosenberger (2003) formalized a three-step approach for the development of optimal RAR procedures. The first step is to mathematically formulate the objectives, such as maximizing the power, and to derive an optimal allocation given these objectives. For example, when comparing two binary responses with success rates p_A and p_B, given the objective of maximizing the power, the optimal allocation proportions η_A and η_B for the two treatments will be the Neyman allocation:

$$\eta_A = \frac{\sqrt{p_A q_A}}{\sqrt{p_A q_A} + \sqrt{p_B q_B}} \text{ and } \eta_B = \frac{\sqrt{p_B q_B}}{\sqrt{p_A q_A} + \sqrt{p_B q_B}},$$

where $q_A = 1 - p_A$ and $q_B = 1 - p_B$. The second step is to develop RAR procedures such as the doubly adaptive biased coin design (DBCD) proposed by Hu and Zhang (2004) to target the theoretically optimal allocation derived in the first step. The third step is to study the operating characteristics of the proposed procedure. It is clear that the derivation of the optimal allocation for certain goals is essential for RAR procedures. In addition, the closed form of the theoretically optimal allocation plays an crucial role in evaluating the properties of the RAR procedure such as the lower bound of the variability Hu et al. (2006), although the numerical results are sufficient to apply the procedure in practice.

A general framework for the optimization problem was given by Jennison and Turnbull (2000), triggering a series of publications on optimal RAR designs with binary outcomes (Rosenberger et al. (2001); Ivanova and Rosenberger (2001); Rosenberger and Hu (2004); Jeon and Hu (2010)), normal outcomes (Biswas and Mandal (2004); Zhang and Rosenberger (2006); Gwise et al. (2008); Biswas and Bhattacharya (2009),), and survival outcomes (Zhang and Rosenberger (2007)). Tymofyeyev et al. (2007) established a general mathematical framework for obtaining optimal allocations using the Karush-Kuhn-Tucker (KKT) conditions (Kuhn and Tucker (1951); Karush (1939)). However, all these studies explored trials with a single endpoint.

We investigate optimal allocation for clinical trials with two possibly correlated endpoints, and we implement this using RAR designs. We study how to maximize the power of a test of homogeneity for a fixed sample size. We first obtain the analytical solution of the optimal allocations for this problem. It is not trivial to check the applicability of the KKT conditions in the presence of multiple endpoints with a covariance structure. This leads to several higher-degree equations that are difficult to solve analytically. The numerical solutions are often sufficient to implement the RAR procedure. However, without the derived closed form, it is difficult to study further theoretical properties of the RAR designs, the third step of the process (Hu and Rosenberger (2003)). This is one of the major contributions of this section. Further, we have performed comprehensive numerical studies based on simulated data and the redesign of a clinical trial. Our method is shown to be efficient than traditional designs. The power is moderately increased using the RAR designs for the optimal allocations. Therefore, both our theoretical and numerical results provide important insight for implementing and further investigating RAR designs in clinical trials with multiple endpoints to increase power.

This chapter is organized as follows. Firstly, we present the framework, define the optimization problems, and derive the optimal allocation proportions. Secondly, we

implement the optimal allocation using the DBCD proposed by Hu and Zhang (2004) and report the results of our numerical studies and proof of Theorems is in the Appendix. Setting $\boldsymbol{\omega} = (1, 1)'$ in problem (2.2) maximizes the power for a fixed sample size.

2.2 Framework, notation, and optimization problems

Consider a randomized clinical trial with two treatment groups and two possibly correlated binary endpoints. Assume that there are n_A patients in treatment A and n_B patients in treatment B at the end of the trial, and let $\mathbf{n} = (n_A, n_B)$. Let (p_{A1}, p_{A2}) and (p_{B1}, p_{B2}) be the success rates of the two endpoints in treatments A and B, respectively, and ρ_A and ρ_B be the correlations between these two endpoints for treatments A and B, respectively.

We wish to test the hypothesis

$$H_0 : \mathbf{p} = \mathbf{0} \quad \text{vs} \quad H_A : \mathbf{p} \neq \mathbf{0}$$

where $\mathbf{p} = (p_{A1} - p_{B1}, p_{A2} - p_{B2})$. Let $\hat{\mathbf{p}} = (\hat{p}_{A1} - \hat{p}_{B1}, \hat{p}_{A2} - \hat{p}_{B2})$ be the maximum likelihood estimator (MLE) of $\mathbf{p}$ and $\hat{\boldsymbol{\Sigma}}$ the estimator of $\boldsymbol{\Sigma} = Var(\hat{\mathbf{p}})$. Here, $\boldsymbol{\Sigma}$ can be written as

$$\boldsymbol{\Sigma} = \begin{bmatrix} \frac{p_{A1}q_{A1}}{n_A} + \frac{p_{B1}q_{B1}}{n_B} & P \\ P & \frac{p_{A2}q_{A2}}{n_A} + \frac{p_{B2}q_{B2}}{n_B} \end{bmatrix},$$

where $q_{A1} = 1 - p_{A1}$, $q_{A2} = 1 - p_{A2}$, $q_{B1} = 1 - p_{B1}$, $q_{B2} = 1 - p_{B2}$, and P is defined to be $\rho_A\sqrt{p_{A1}q_{A1}p_{A2}q_{A2}}/n_A + \rho_B\sqrt{p_{B1}q_{B1}p_{B2}q_{B2}}/n_B$, the covariance of $(\hat{p}_{A1} - \hat{p}_{B1})$ and $(\hat{p}_{A2} - \hat{p}_{B2})$.

We use $\hat{\mathbf{p}}\hat{\boldsymbol{\Sigma}}^{-1}\hat{\mathbf{p}}'$ as the test statistic; it converges to χ^2_2 under the null hypothesis. Under the alternative hypothesis, it follows a noncentral chi-squared distribution $\chi^2_2(\phi(\mathbf{n}))$

with noncentrality parameter $\phi(\mathbf{n})$ defined as

$$\begin{aligned}\phi(\mathbf{n}) &= \mathbf{p}\boldsymbol{\Sigma}^{-\mathbf{1}}\mathbf{p}^{'} \\ &= \left\{(p_{A1}-p_{B1})^2\left(\frac{p_{A2}q_{A2}}{n_A}+\frac{p_{B2}q_{B2}}{n_B}\right)+(p_{A2}-p_{B2})^2\left(\frac{p_{A1}q_{A1}}{n_A}+\frac{p_{B1}q_{B1}}{n_B}\right)\right. \\ &\left.-2P(p_{A1}-p_{B1})(p_{A2}-p_{B2})\right\}\Big/\left\{\left(\frac{p_{A1}q_{A1}}{n_A}+\frac{p_{B1}q_{B1}}{n_B}\right)\left(\frac{p_{A2}q_{A2}}{n_A}+\frac{p_{B2}q_{B2}}{n_B}\right)-P^2\right\}.\end{aligned} \tag{2.1}$$

Note that the power can be expressed as an increasing function of the noncentrality parameter $\phi(\mathbf{n})$ Patnaik (1949).

We now formulate the optimization problems as to maximizes the noncentrality parameter of the test under certain conditions on the weighted sum of the sample sizes:

$$\begin{aligned}&\max_{n_A,n_B}\ \phi(n_A,n_B) \\ &s.t.\ \ n_A,n_B \geq D(n_A+n_B),\ \omega_A n_A+\omega_B n_B \leq M.\end{aligned} \tag{2.2}$$

Here M is a positive constant, the weight ω_A and ω_B are also positive, and the constant $D \in [0, 1/2]$ is the lower bound for the allocation proportion $n_A/(n_A+n_B)$ and $n_B/(n_A+n_B)$, that allows us to control the feasible region and avoid having too few people in either of the treatments. Our goal is to find the allocation proportion, $\eta_A^* = n_A^*/(n_A^*+n_B^*)$ and $\eta_B^* = 1-\eta_A^*$, as the solution of the optimization problem. Here, the positive M is used to define the optimization problem, and our final solutions with regard to the allocation proportions instead of sample size do not depend on M. The value of D is less important for clinical trials with two treatments than those with multiple treatments where $D = 0$ may lead to solutions on the boundary that involves only the best and the worst treatments. We prove that the noncentrality parameter $\phi(n_A, n_B)$ in (1) is a concave function, and $\nabla\phi \geq 0$ in the Appendix.

2.3 Karush–Kuhn–Tucker conditions

The Karush–Kuhn–Tucker (KKT) conditions are used to find the solution for optimization problems constrained to one or more equalities and inequalities. Consider the nonlinearly constrained program

$$\begin{aligned} &\max f(x) \\ &g(x) \leq 0, h(x) = 0 \end{aligned} \tag{2.3}$$

where f, g and h are continuously differentiable function from $\mathbb{R}^n$ to $\mathbb{R}, \mathbb{R}^p$ and $\mathbb{R}^p$ respectively. The KKT system for this problem is:

$$\begin{aligned} &\nabla f(x) + \sum_{j=1}^{p} u_j \nabla g_j(x) + \sum_{j=1}^{q} v_j \nabla h_j(x) = 0, \\ &u \geq 0, g(x) \leq 0, U^T g(x) = 0 \\ &h(x) = 0 \end{aligned} \tag{2.4}$$

where $N = n + p + q$

2.4 Optimized allocation to maximize the power for a fixed sample size

We first consider the optimal allocation to maximize the power for a fixed sample size.

Theorem 1 Let $r = \eta_A^*/\eta_B^*$ and $\tilde{D} = D/(1-D)$. Then the solution for problem (2) with weights $\boldsymbol{\omega} = (\omega_A, \omega_B)' = (1,1)'$ is as follows:

(1) If $p_{A1} = p_{B1}, p_{A2} = p_{B2}$, η_A^* can be any number between 0 and 1 greater than D, and r can be any number between $\tilde{D}$ and $\tilde{D}^{-1}$.

(2) If $(p_{A1}-p_{B1})^2+(p_{A2}-p_{B2})^2\neq 0$, let

$$H=b^2-ac,\ I=ae-4bd+3c^2,\ G=a^2d-3abc+2b^3,\ J=\frac{4H^3-a^2HI-G^2}{a^3}$$

where

$$\begin{aligned}a=&(p_{A1}-p_{B1})^2p_{B1}q_{B1}p_{B2}^2q_{B2}^2(1-\rho_B^2)+(p_{A2}-p_{B2})^2p_{B1}^2q_{B1}^2p_{B2}q_{B2}(1-\rho_B^2)\\&-2(p_{A1}-p_{B1})(p_{A2}-p_{B2})(p_{B1}q_{B1}p_{B2}q_{B2})^{\frac{3}{2}}\rho_B(1-\rho_B^2)\end{aligned}$$

$$\begin{aligned}b=&\frac{1}{2}(p_{A1}-p_{B1})^2p_{A2}q_{A2}p_{B1}q_{B1}p_{B2}q_{B2}(1-\rho_B^2)\\&+\frac{1}{2}(p_{A2}-p_{B2})^2p_{A1}q_{A1}p_{B1}q_{B1}p_{B2}q_{B2}(1-\rho_B^2)\\&-(p_{A1}-p_{B1})(p_{A2}-p_{B2})\sqrt{p_{A1}q_{A1}p_{A2}q_{A2}}p_{B1}q_{B1}p_{B2}q_{B2}\rho_A(1-\rho_B^2)\end{aligned}$$

$$
\begin{aligned}
c = \Big[& (p_{A1}-p_{B1})^2 \left\{ p_{A2}^2 q_{A2}^2 p_{B1} q_{B1} - p_{A1} q_{A1} p_{B2}^2 q_{B2}^2 + p_{A1} q_{A1} p_{A2} q_{A2} p_{B2} q_{B2} \rho_A^2 \right. \\
& - p_{A2} q_{A2} p_{B1} q_{B1} p_{B2} q_{B2} \rho_B^2 \\
& \left. - 2(p_{A2} q_{A2} - p_{B2} q_{B2}) \sqrt{p_{A1} q_{A1} p_{A2} q_{A2} p_{B1} q_{B1} p_{B2} q_{B2}} \rho_A \rho_B \right\} \\
& + (p_{A2}-p_{B2})^2 \left\{ p_{A1}^2 q_{A1}^2 p_{B2} q_{B2} - p_{A2} q_{A2} p_{B1}^2 q_{B1}^2 + p_{A1} q_{A1} p_{A2} q_{A2} p_{B1} q_{B1} \rho_A^2 \right. \\
& - p_{A1} q_{A1} p_{B1} q_{B1} p_{B2} q_{B2} \rho_B^2 \\
& \left. - 2(p_{A1} q_{A1} - p_{B1} q_{B1}) \sqrt{p_{A1} q_{A1} p_{A2} q_{A2} p_{B1} q_{B1} p_{B2} q_{B2}} \rho_A \rho_B \right\} \\
& - 2(p_{A1}-p_{B1})(p_{A2}-p_{B2}) \times \\
& \left\{ (p_{A1} q_{A1} p_{B2} q_{B2} + p_{A2} q_{A2} p_{B1} q_{B1} + p_{B1} q_{B1} p_{B2} q_{B2}) \sqrt{p_{A1} q_{A1} p_{A2} q_{A2}} \rho_A \right. \\
& - (p_{A1} q_{A1} p_{B2} q_{B2} + p_{A2} q_{A2} p_{B1} q_{B1} + p_{A1} q_{A1} p_{A2} q_{A2}) \sqrt{p_{B1} q_{B1} p_{B2} q_{B2}} \rho_B \\
& \left. - p_{A1} q_{A1} p_{A2} q_{A2} \sqrt{p_{B1} q_{B1} p_{B2} q_{B2}} \rho_A^2 \rho_B + p_{B1} q_{B1} p_{B2} q_{B2} \sqrt{p_{A1} q_{A1} p_{A2} q_{A2}} \rho_A \rho_B^2 \right\} \Big] / 6
\end{aligned}
$$

$$
\begin{aligned}
d = & -\frac{1}{2} (p_{A1}-p_{B1})^2 p_{A1} q_{A1} p_{A2} q_{A2} p_{B2} q_{B2} (1-\rho_A^2) \\
& - \frac{1}{2} (p_{A2}-p_{B2})^2 p_{A1} q_{A1} p_{A2} q_{A2} p_{B1} q_{B1} (1-\rho_A^2) \\
& + (p_{A1}-p_{B1})(p_{A2}-p_{B2}) p_{A1} q_{A1} p_{A2} q_{A2} \sqrt{p_{B1} q_{B1} p_{B2} q_{B2}} \rho_B (1-\rho_A^2)
\end{aligned}
$$

$$
\begin{aligned}
e = & -(p_{A1}-p_{B1})^2 p_{A1} q_{A1} p_{A2}^2 q_{A2}^2 (1-\rho_A^2) \\
& - (p_{A2}-p_{B2})^2 p_{A1}^2 q_{A1}^2 p_{A2} q_{A2} (1-\rho_A^2) \\
& + 2(p_{A1}-p_{B1})(p_{A2}-p_{B2}) (p_{A1} q_{A1} p_{A2} q_{A2})^{\frac{3}{2}} \rho_A (1-\rho_A^2).
\end{aligned}
$$

(i) If $G = 0$, $r = \max\left((-b + \sqrt{3H + \sqrt{12H^2 - a^2 I}})/a, \tilde{D}\right)$.

(ii) If $G \neq 0$, $I = J = 0$, $r = \max\left(\left\{-b - 3sign(G)\sqrt{H}\right\}/a, \tilde{D}\right)$.

(iii) Let $\triangle = I^3 - 27J^2$. If $G \neq 0$, $I^2 + J^2 \neq 0$, $\Delta < 0$, $r = \max\big((-b - sign(G)\sqrt{t} + \sqrt{|G|/\sqrt{t} - t + 3H})/a, \tilde{D}\big)$.

(iv) If $G \neq 0$, $I^2 + J^2 \neq 0, \Delta \geq 0$, $r = \max\left((-b + \sqrt{y_1} + \sqrt{y_2} + \sqrt{y_3})/a, \tilde{D}\right)$, where

$$y_1 = a\sqrt{\frac{|I|}{3}} cos\left(\frac{\theta}{3}\right) + H,\ y_{2,3} = a\sqrt{\frac{|I|}{3}} cos\left(\frac{\theta}{3} \pm \frac{2\pi}{3}\right) + H,$$

$$\theta = cos^{-1}\left(\frac{-J}{\sqrt{|I|^3/27}}\right).$$

Remark 1. When there is no difference between the two groups for one endpoint and no correlation between the endpoints, the allocation reduces to the Neyman allocation. That is, when $p_{A1} = p_{B1}, p_{A2} \neq p_{B2}$, and $\rho_A = \rho_B = 0$, we have $r = \max(\sqrt{p_{A2}q_{A2}}/\sqrt{p_{B2}q_{B2}}, \tilde{D})$. Because we assign patients to the different treatments based on just one endpoint it is natural that we get the same result as for the one-endpoint problem.

2.5 Implementing the optimal allocation using response-adaptive randomization

In the literature, a variety of randomization approaches such as the sequential maximum likelihood (SMLE) (Melfi *et al.*, 2001) have been proposed for optimal allocation procedures. We implement the optimal allocations in Theorems 1 and 2 using the DBCD proposed by Hu and Zhang (2004), since its advantages are widely accepted. Assuming that the responses are available immediately, the DBCD can be described as follows:

1. We assign the first $2a$ subjects (usually 10%–20% of the total sample size) to the two treatment groups by certain restricted randomization procedures such as permuted block designs (Rosenberger and Lachin (2002))

2. After $j-1$ ($j > 2a$) subjects have been assigned, we estimate the unknown parameters based on the $j-1$ observations, and we calculate the estimated optimal allocations based on Theorem 1 or 2.

3. We assign the jth subject to treatment A with probability ψ,

$$\psi(v_A, \hat{\eta}_A^*) = \begin{cases} 1 & \text{if } v_A = 0 \\ \frac{\hat{\eta}_A^*(\hat{\eta}_A^*/v_A)^\gamma}{\hat{\eta}_A^*(\hat{\eta}_A^*/v_A)^\gamma + (1-\hat{\eta}_A^*)((1-\hat{\eta}_A^*)/(1-v_A))^\gamma} & \text{if } 0 < v_A < 1\,, \\ 0 & \text{if } v_A = 1 \end{cases} \tag{2.5}$$

where v_A is the currently observed allocation proportion of the patients in treatment A, $\hat{\eta}_A^*$ is the estimated target allocation proportion, and γ is a tuning parameter that adjusts the balance between the extent of randomization and the convergence rates of the allocation proportion to the targets. In our numerical studies, the first 10% of the patients are allocated to the two treatments using restricted randomization to obtain initial parameter estimates, and then we switch to the DBCD procedure with $\gamma = 2$.There are few theoretical studies about the choice of the tuning parameters in the literature, though they are popular for adaptive randomization designs. Numerical studies showed that $\gamma = 2$ is a good trade-off (Rosenberger and Hu (2004)), and this value has been widely used in the literature (Tymofyeyev et al. (2007); Zhu et al. (2010), Zhu and Hu (2012)). All the results are based on 5000 replications.

2.6 Simulation studies

First we study the type I error rate with $p_{A1} = p_{B1} = p_{A2} = p_{B2} = 0.3$. 500 patients are sequentially enrolled in the trial. We assign 7 different sets of weights with relative weights (w_1/w_2) range from 0.1 to 10, and the corresponding importance of endpoint 2 decreases as the relative weights increase. In Table 3.4.1 we report the type I error rates (α) for different correlations between the two endpoints when DBCD is implemented targeting the optimal allocation proportions obtained in Theorems 1. We can see that type I error rate can be well controlled. In addition, the average parameter estimates and the corresponding standard deviations out of 5000 replications are also reported. Even if the RAR design involves complicated sequential updating procedures, we can still estimate the parameters very well.

We consider the following scenarios to study the efficient and ethical advantages of RAR procedures targeting the derived optimal allocation proportions in this paper:

Scenario 1: Treatment B is better than treatment A in terms of both endpoints, and the success rates are $p_{A1} = 0.20$, $p_{B1} = 0.35$, $p_{A2} = 0.30$, and $p_{B2} = 0.45$.

Scenario 2: The treatment effects of the two groups are the same in terms of endpoint 1, and treatment A is better than treatment B in terms of endpoint 2. The success rates are $p_{A1} = 0.25$, $p_{B1} = 0.25$, $p_{A2} = 0.45$, and $p_{B2} = 0.30$.

The above parameters are chosen to show the advantages of our procedure when the treatment effects are relatively conservative. RAR designs targeting the optimal allocation proportions may perform better for other parameter combinations.

In Table 2.6.1, we compare complete randomization (CR) with DBCD targeting the optimal allocation proportion obtained in Theorem 1, where power is the focus. The sample sizes (N) are chosen in each case to make the power to be about 0.8. We can see that our method consistently increases the power, which could lead to a reduction in the

sample size in real trials. Table 2.6.1 also shows that the correlations have a dramatic influence on the sample size needed to achieve a given level of power.

Based on the asymptotic properties derived by Hu and Zhang (2004), the type I error rate for the DBCD can be easily controlled, as shown in Table 3.4.1. But it is worth studying the performance of our procedure in terms of type I error rate when the sample size is small. In Table 2.6.2, we report results for certain sample sizes in Table 2.6.1 when our procedure is targeting the optimal allocation proportions in Theorem 1 and $p_{A1} = p_{B1} = p_{A2} = p_{B2} = 0.3$. We find that the type I error rate is not quite satisfactory when the sample size is less than 250. This could be due to the convergence rate of the RAR procedures.

Table 2.6.1: Comparison of complete randomization and DBCD targeting the optimal allocation in Theorem 1

ρ_A	ρ_B	N	Power	
			CR	RAR
$p_{A1} = 0.20$	$p_{B1} = 0.35$	$p_{A2} = 0.30$	$p_{B2} = 0.45$	
0.0	0.0	178	0.794	0.808
0.1	0.1	197	0.798	0.811
0.3	0.1	210	0.800	0.809
0.3	0.3	234	0.803	0.809
0.5	0.3	250	0.799	0.805
$p_{A1} = 0.25$	$p_{B1} = 0.25$	$p_{A2} = 0.45$	$p_{B2} = 0.30$	
0.0	0.0	395	0.798	0.809
0.1	0.1	395	0.802	0.807
0.3	0.1	380	0.800	0.811
0.3	0.3	360	0.801	0.807
0.5	0.3	334	0.798	0.811

Table 2.6.2: Type I error rates and parameter estimates for RAR design targeting optimal allocations in Theorems 1 when sample sizes are smaller

N	ρ_A	ρ_B	$\hat{p}_{A1}$	$\hat{p}_{A2}$	$\hat{p}_{B1}$	$\hat{p}_{B2}$	$\hat{\rho}_A$	$\hat{\rho}_B$	α
178	0	0	0.299(0.050)	0.299(0.048)	0.299(0.049)	0.300(0.048)	0.007(0.107)	0.008(0.103)	0.060
234	0.3	0.3	0.299(0.043)	0.299(0.042)	0.299(0.043)	0.298(0.042)	0.300(0.098)	0.300(0.095)	0.062
250	0.5	0.3	0.299(0.042)	0.299(0.041)	0.299(0.042)	0.298(0.040)	0.502(0.088)	0.298(0.091)	0.057
334	0.5	0.3	0.299(0.036)	0.299(0.035)	0.300(0.036)	0.299(0.034)	0.502(0.076)	0.300(0.078)	0.053
360	0.3	0.3	0.299(0.034)	0.299(0.034)	0.300(0.035)	0.299(0.033)	0.300(0.077)	0.301(0.077)	0.052
395	0	0	0.300(0.033)	0.300(0.032)	0.300(0.033)	0.299(0.032)	0.002(0.072)	0.002(0.069)	0.049

2.7 Re-designing the trial of perioperative total parenteral nutrition (TPN) treatment using the RAR design

Tang et al. (1989a) discussed a trial comparing perioperative TPN treatment with no perioperative TPN in nutritionally compromised patients about to undergo surgery for gastric cancer. The investigators were interested in whether the perioperative TPN reduced the proportion of major and minor complications in the week following surgery. Each patient was at risk of major and minor complications that were well defined and mutually exclusive. Hence, these two endpoints were naturally correlated. Tang et al. (1989a) carried out a numerical study with the following parameter combinations: $p_{A1} = 0.25, p_{B1} = 0.15, p_{A2} = 0.45$, and $p_{B2} = 0.30$. We redesign the trial and generate datasets using the same parameters. In Table 2.7.1, we report the results of DBCD targeting the optimal allocation obtained from Theorem 1 that aims to maximize the power, and we compare this procedure with CR. In Table 3.5.1, we report the results of DBCD targeting the optimal allocation obtained from Theorem 2 that aims to minimize the EWNF while maintaining the power, and we compare it with CR. In each scenario, we consider the correlations between the endpoints to range from 0 to 0.5, and we adjust the total sample size to achieve 80% power for CR. Table 2.7.1 shows that the RAR design

can either achieve higher power of detecting the treatment effects under different levels of correlation between the endpoints, or reduce the number of failures while holding the power to the same level as that for CR.

Table 2.7.1: Comparison of complete randomization and DBCD targeting the optimal allocation in Theorem 1 for redesigned trial

ρ_A	ρ_B	N	Power	
			CR	RAR
0	0	240	0.802	0.818
0.1	0.1	260	0.799	0.807
0.2	0.2	285	0.801	0.812
0.3	0.1	280	0.796	0.802
0.3	0.3	300	0.799	0.802
0.5	0.3	335	0.795	0.813
0.5	0.5	349	0.797	0.805

2.8 Discussion

Multiple endpoints are widely employed in modern clinical trials. In patients with coronary heart disease, we may be interested in both resting and exercise ejection fractions. In blood-pressure trials, we might be interested in diastolic and systolic blood pressure or mean arterial pressure and pulse pressure. In stroke treatment a number of scales are used to measure recovery, and no one scale is believed to assess all the dimensions. However, there has been little study of how to ensure that such trials are ethical and efficient through well-designed randomization procedures. In this paper, we use RAR designs to implement optimal allocation in order to maximize the power. We give the analytical solutions to two general optimization problems for nonlinear models by using the KKT conditions. We have discussed the theoretical results and the advantages of our method in the previous sections. We now discuss possible concerns and directions

for future research.

In general, adaptive designs may need a larger sample size than that for CR because of the larger variation in the allocation proportions. However, the traditional formula for the CR sample size also ignores the variation of the randomization procedure, so the actual loss of power under adaptive designs is less significant. This paper shows that targeting the appropriate allocation proportion with RAR designs can often increase the power and reduce the required sample size. The impression that equal allocation results in the best power is not always correct.

Clinical trials are complex and have different types of clinical decision rules, inferential goals, and common approaches to analyze multiple endpoints for clinical trials (Dmitrienko et al. (2009)), among which one important rule is that the trial's outcome would be declared positive if at least one endpoint was significantly improved compared to the control. One motivating example is the STRIVE study (Zeiher et al. (2004)) which aimed to determine whether sivelestat would reduce 28-day all-cause mortality or increase the number of ventilator-free days compared with placebo in mechanically ventilated patients with acute lung injury. This paper focuses on this scenario. It is worth noting that there are other clinical decision rules. In some trials, a clinically meaningful effect is defined as the simultaneous improvement in multiple endpoints, and multiple one-sided alternative hypotheses will be tested. In the Introduction section, we introduced the three-step approach for the development of optimal RAR procedures formalized by Hu and Rosenberger (2003). If the objectives are changed and the corresponding optimization problems such as equation (2.2) is changed, the theoretically optimal allocation proportions will usually be different.

Chapter 3

Implementing Optimal Allocation in Clinical Trials with Multiple Endpoints to Minimize the expected number of treatment failures for a fixed power

3.1 Introduction

Randomized controlled trials often considered the gold standard for clinical trials. Objects are randomly assigned to control and treatment groups to reduce selection bias and allocation bias. However ethical issues are raised when objects are assigned to treatments for which there already partial evidence of its inferiority. That's the reason why trials are often monitored an interim basis.

The COVID-19 pandemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) hit the world with more than 11 million cases and more than 500,000 deaths. The fatality rate is 4.6% worldwide and can be as high as 13.3% as observed in the Italian population as of April 23rd 2020(Quintaliani et al. (2020)). Many countries started clinical trials of potential treatments such as hydroxychloroquine and azithromycin. With such high mortality rate it's more ethical to assign patients to treatments showing higher potential from early data to increase the number of successes. In the study design to

evaluate the role of crystalloid preload in the prevention of hypotension associated with spinal anesthesia for elective cesarean section(Rout et al. (1993)), the criterion was to expose the minimum number of patients to management that tend to cause harm. Another example can be find in Palmer (1991), the treatment selection criterion was to maximize expected numbers of future successes which is equavalent. RAR designs can also be helpful when multiple agents and combinations can be investigated simultaneously. In which case, as the clinical trial progresses the randomization ratio will favor the treatments that are performing better. These examples motivate our investigation of the use of RAR designs to find optimal allocations for clinical trials with multiple endpoints to achieve ethical objectives.

This chapter is organized as follows. Firstly, we present the framework, define the optimization problems to minimize the number of failures, and derive the optimal allocation proportions. Secondly, we implement the optimal allocation using the DBCD proposed by Hu and Zhang (2004) and report the results of our numerical studies.

3.2 Framework, notation, and optimization problems

Under the same setting and notation as described in Section 2.2 considering a randomized clinical trial with two treatment groups and two possibly correlated binary endpoints. To test the hypothesis

$$H_0 : \mathbf{p} = \mathbf{0} \quad \text{vs} \quad H_A : \mathbf{p} \neq \mathbf{0}$$

where $\mathbf{p} = (p_{A1} - p_{B1}, p_{A2} - p_{B2})$. we use $\hat{\mathbf{p}}\hat{\mathbf{\Sigma}}^{-1}\hat{\mathbf{p}}'$ as the test statistic; it converges to χ^2_2 under the null hypothesis. Under the alternative hypothesis, it follows a noncentral

chi-squared distribution $\chi^2_2(\phi(\mathbf{n}))$ with noncentrality parameter $\phi(\mathbf{n})$ defined as

$$\begin{aligned}\phi(\mathbf{n}) &= \mathbf{p}\boldsymbol{\Sigma}^{-\mathbf{1}}\mathbf{p}^{'} \\ &= \left\{(p_{A1}-p_{B1})^2\left(\frac{p_{A2}q_{A2}}{n_A}+\frac{p_{B2}q_{B2}}{n_B}\right)+(p_{A2}-p_{B2})^2\left(\frac{p_{A1}q_{A1}}{n_A}+\frac{p_{B1}q_{B1}}{n_B}\right)\right. \\ &\left.-2P(p_{A1}-p_{B1})(p_{A2}-p_{B2})\right\}\Big/\left\{\left(\frac{p_{A1}q_{A1}}{n_A}+\frac{p_{B1}q_{B1}}{n_B}\right)\left(\frac{p_{A2}q_{A2}}{n_A}+\frac{p_{B2}q_{B2}}{n_B}\right)-P^2\right\}\end{aligned} \tag{3.1}$$

The optimization problem becomes to minimize the weighted sum of the sample sizes under certain conditions on the noncentrality parameter:

$$\begin{aligned}&\min_{n_A,n_B}\ \omega_A n_A+\omega_B n_B \\ &s.t.\ \ n_A,n_B\geq D(n_A+n_B),\ \phi(n_A,n_B)\geq C\end{aligned} \tag{3.2}$$

where C is a positive constant and D is the constant introduced above. Here, the positive C are used to define the optimization problem, and our final solutions with regard to the allocation proportions instead of sample size do not depend on C.

The equivalence of this optimization problem and the one discussed in first chapter with regard to the same allocation proportions was discussed by Tymofyeyev et al. (2007). We can vary the optimization problems by changing the weights $\boldsymbol{\omega}=(\omega_A,\omega_B)^{'}$. Setting $\boldsymbol{\omega}=(q_{A1}+q_{A2},q_{B1}+q_{B2})'$ in problem (3.2) minimizes the EWNF for a fixed power.

3.3 Optimized allocation to minimize the expected number of treatment failures for a fixed power

We consider the optimal allocation to minimize the EWNF for a fixed power. We assign different weights to the two endpoints to emphasize the importance of the individual

endpoint.

Let $w_j \geq 0$, $j = 1, 2$, be the weight for endpoint j. Then the EWNF is

$$w_1 q_{A1} n_A + w_1 q_{B1} n_B + w_2 q_{A2} n_A + w_2 q_{B2} n_B$$
$$= (w_1 q_{A1} + w_2 q_{A2}) n_A + (w_1 q_{B1} + w_2 q_{B2}) n_B.$$

This is problem (3.2) with $\boldsymbol{\omega} = (w_1 q_{A1} + w_2 q_{A2}, w_1 q_{B1} + w_2 q_{B2})'$. When $w_1 = w_2 = 1$, the two endpoints are equally important.

Theorem 2 Let $r = \eta_A^*/\eta_B^*$ and $\tilde{D} = D/(1-D)$. Then the solution for problem (3.2) with $\boldsymbol{\omega} = (w_1 q_{A1} + w_2 q_{A2}, w_1 q_{B1} + w_2 q_{B2})'$ is as follows:

(1) If $p_{A1} = p_{B1}, p_{A2} = p_{B2}$, η_A^* can be any number between 0 and 1 greater than D, and r can be any number between $\tilde{D}$ and $\tilde{D}^{-1}$.

(2) If $(p_{A1} - p_{B1})^2 + (p_{A2} - p_{B2})^2 \neq 0$, let $\omega_A = w_1 q_{A1} + w_2 q_{A2}$ and $\omega_B = w_1 q_{B1} + w_2 q_{B2}$. Then the optimal allocations have the same form as in Theorem 1 with

$$\begin{aligned} a =& \omega_A (p_{A1} - p_{B1})^2 p_{B1} q_{B1} p_{B2}^2 q_{B2}^2 (1 - \rho_B^2) \\ &+ \omega_A (p_{A2} - p_{B2})^2 p_{B1}^2 q_{B1}^2 p_{B2} q_{B2} (1 - \rho_B^2) \\ &- 2\omega_A (p_{A1} - p_{B1})(p_{A2} - p_{B2})(p_{B1} q_{B1} p_{B2} q_{B2})^{\frac{3}{2}} \rho_B (1 - \rho_B^2) \end{aligned}$$

$$\begin{aligned} b =& \frac{1}{2} \omega_A (p_{A1} - p_{B1})^2 p_{B1} q_{B1} p_{A2} q_{A2} p_{B2} q_{B2} (1 - \rho_B^2) \\ &+ \frac{1}{2} \omega_A (p_{A2} - p_{B2})^2 p_{B1} q_{B1} p_{A1} q_{A1} p_{B2} q_{B2} (1 - \rho_B^2) \\ &- (p_{A1} - p_{B1})(p_{A2} - p_{B2}) p_{B1} q_{B1} p_{B2} q_{B2} \sqrt{p_{A1} q_{A1} p_{A2} q_{A2}} \rho_A \omega_A (1 - \rho_B^2) \end{aligned}$$

$$
\begin{aligned}
c =& \Big((p_{A1}-p_{B1})^2 \left\{ -\omega_B p_{A1} q_{A1} p_{B2}^2 q_{B2}^2 + \omega_A p_{B1} q_{B1} p_{A2}^2 q_{A2}^2 + p_{A1} q_{A1} p_{A2} q_{A2} p_{B2} q_{B2} \rho_A^2 \omega_A \right. \\
& - p_{A2} q_{A2} p_{B1} q_{B1} p_{B2} q_{B2} \rho_B^2 \omega_B \\
& \left. - 2(p_{A2} q_{A2} \omega_A - p_{B2} q_{B2} \omega_B) \sqrt{p_{A1} q_{A1} p_{A2} q_{A2} p_{B1} q_{B1} p_{B2} q_{B2}} \rho_A \rho_B \right\} \\
& + (p_{A2}-p_{B2})^2 \left\{ -\omega_B p_{A2} q_{A2} p_{B1}^2 q_{B1}^2 + \omega_A p_{B2} q_{B2} p_{A1}^2 q_{A1}^2 + p_{A1} q_{A1} p_{A2} q_{A2} p_{B1} q_{B1} \rho_A^2 \omega_A \right. \\
& - p_{A1} q_{A1} p_{B1} q_{B1} p_{B2} q_{B2} \rho_B^2 \omega_B \\
& \left. - 2(p_{A1} q_{A1} \omega_A - p_{B1} q_{B1} \omega_B) \sqrt{p_{A1} q_{A1} p_{A2} q_{A2} p_{B1} q_{B1} p_{B2} q_{B2}} \rho_A \rho_B \right\} \\
& - 2 \Big[\Big\{ (p_{A2} q_{A2} p_{B1} q_{B1} + p_{A1} q_{A1} p_{B2} q_{B2}) \omega_A \rho_A + p_{B1} q_{B1} p_{B2} q_{B2} \omega_B \rho_A (1+\rho_B^2) \Big\} \times \\
& \sqrt{p_{A1} q_{A1} p_{A2} q_{A2}} \\
& - \Big\{ (p_{A2} q_{A2} p_{B1} q_{B1} + p_{A1} q_{A1} p_{B2} q_{B2}) \omega_B \rho_B + p_{A1} q_{A1} p_{A2} q_{A2} \omega_A \rho_B (1+\rho_A^2) \Big\} \times \\
& \sqrt{p_{B1} q_{B1} p_{B2} q_{B2}} \Big] (p_{A1}-p_{B1})(p_{A2}-p_{B2}) \Big) \Big/ 6
\end{aligned}
$$

$$
\begin{aligned}
d =& -\frac{1}{2} \omega_B (p_{A1}-p_{B1})^2 p_{A1} q_{A1} p_{A2} q_{A2} p_{B2} q_{B2} (1-\rho_A^2) \\
& - \frac{1}{2} \omega_B (p_{A2}-p_{B2})^2 p_{A1} q_{A1} p_{A2} q_{A2} p_{B1} q_{B1} (1-\rho_A^2) \\
& + (p_{A1}-p_{B1})(p_{A2}-p_{B2}) p_{A1} q_{A1} p_{A2} q_{A2} \sqrt{p_{B1} q_{B1} p_{B2} q_{B2}} \rho_B \omega_B (1-\rho_A^2)
\end{aligned}
$$

$$
\begin{aligned}
e =& -\omega_B (p_{A1}-p_{B1})^2 p_{A1} q_{A1} p_{A2}^2 q_{A2}^2 (1-\rho_A^2) \\
& - \omega_B (p_{A2}-p_{B2})^2 p_{A1}^2 q_{A1}^2 p_{A2} q_{A2} (1-\rho_A^2) \\
& + 2(p_{A1}-p_{B1})(p_{A2}-p_{B2}) (p_{A1} q_{A1} p_{A2} q_{A2})^{\frac{3}{2}} \rho_A \omega_B (1-\rho_A^2).
\end{aligned}
$$

Remark 2. One special case is that when $p_{A1} = p_{B1}, p_{A2} \neq p_{B2}, \rho_A = \rho_B = 0$, the expected weighted number of failures is:

$$
w_1 q_{A1} n_A + w_1 q_{B1} n_B + w_2 q_{A2} n_A + w_2 q_{B2} n_B = w_2 (q_{A2} n_A + q_{B2} n_B) + w_1 q_{A1} n
$$

We only need to minimize $q_{A2}n_A + q_{B2}n_B$, then $\boldsymbol{\omega} = (q_{A2}, q_{B2})'$. We have

$$r = \max(\sqrt{p_{A2}}/\sqrt{p_{B2}}, \tilde{D}).$$

The above special case reduces to the optimal allocation proposed by Rosenberger et al. (2001), when there is no difference between the two groups for one endpoint and no correlation between the endpoints. This is because the optimal allocations only depend on one endpoint.

Remark 3. Although the theoretical solutions of the optimal allocations are complicated, we can easily obtain these results with a low computing cost since they are functions of the estimated parameters.

3.4 Simulation studies

We consider the similar setting as described in Section 2.6

In Table 3.4.2, we compare CR with DBCD targeting the optimal allocation proportion obtained in Theorem 2, and we report the EWNF and the power. Same weights as in Table 3.4.1 are assigned to those two endpoints. Table 3.4.2 shows that our procedures reduced the EWNF for all the cases while maintaining the same level of power. The advantages can be seen even in Scenario 2. Table 3.4.2 also shows that the correlations have a dramatic influence on the sample size needed to achieve a given level of power which echos the finding in 2.6. Simulations with small sample size is small are also performed. In Table 2.6.2, we report results for certain sample sizes in Table 2.6.1 when our procedure is targeting the optimal allocation proportions in Theorem 1 and $p_{A1} = p_{B1} = p_{A2} = p_{B2} = 0.3$. We find that the type I error rate is not quite satisfactory when the sample size is less than 250. The results are quite similar when our procedure is targeting the optimal allocation proportions in Theorem 2.

Table 3.4.1: Type I error rates and parameter estimates for DBCD targeting optimal allocations in Theorems 1 and 2

ρ_A	ρ_B	w_1	w_2	$\hat{p}_{A1}$	$\hat{p}_{A2}$	$\hat{p}_{B1}$	$\hat{p}_{B2}$	$\hat{\rho}_A$	$\hat{\rho}_B$	α
Targeting optimal allocations in Theorem 2										
0	0	1	0.1	0.300(0.029)	0.299(0.029)	0.300(0.029)	0.299(0.029)	0.001(0.063)	0.001(0.063)	0.051
0	0	1	0.25	0.299(0.029)	0.300(0.029)	0.300(0.029)	0.300(0.030)	0.000(0.063)	0.002(0.063)	0.052
0	0	1	0.5	0.300(0.029)	0.299(0.029)	0.299(0.029)	0.300(0.029)	0.000(0.063)	0.000(0.064)	0.053
0	0	1	1	0.299(0.029)	0.300(0.029)	0.299(0.029)	0.299(0.029)	0.000(0.063)	0.000(0.063)	0.051
0	0	1	2	0.300(0.029)	0.300(0.029)	0.300(0.029)	0.299(0.029)	-0.001(0.063)	-0.001(0.063)	0.051
0	0	1	4	0.300(0.029)	0.300(0.029)	0.300(0.029)	0.299(0.029)	0.000(0.063)	-0.001(0.063)	0.053
0	0	1	10	0.300(0.029)	0.300(0.029)	0.300(0.029)	0.299(0.029)	0.000(0.063)	0.001(0.063)	0.049
0.3	0.3	1	0.1	0.300(0.029)	0.299(0.029)	0.300(0.029)	0.299(0.029)	0.300(0.065)	0.300(0.065)	0.049
0.3	0.3	1	0.25	0.299(0.028)	0.300(0.029)	0.300(0.029)	0.300(0.030)	0.300(0.065)	0.301(0.065)	0.053
0.3	0.3	1	0.5	0.299(0.029)	0.300(0.029)	0.300(0.029)	0.300(0.029)	0.300(0.065)	0.301(0.066)	0.049
0.3	0.3	1	1	0.299(0.028)	0.300(0.029)	0.299(0.029)	0.299(0.029)	0.301(0.065)	0.300(0.066)	0.052
0.3	0.3	1	2	0.300(0.029)	0.300(0.029)	0.300(0.028)	0.299(0.029)	0.299(0.064)	0.298(0.065)	0.052
0.3	0.3	1	4	0.300(0.029)	0.300(0.029)	0.300(0.028)	0.299(0.029)	0.299(0.063)	0.298(0.063)	0.052
0.3	0.3	1	10	0.300(0.029)	0.300(0.029)	0.300(0.029)	0.299(0.029)	0.298(0.064)	0.301(0.066)	0.052
0.5	0.3	1	0.1	0.299(0.029)	0.299(0.029)	0.300(0.029)	0.299(0.029)	0.501(0.060)	0.300(0.065)	0.050
0.5	0.3	1	0.25	0.300(0.029)	0.300(0.029)	0.299(0.029)	0.300(0.030)	0.499(0.061)	0.300(0.065)	0.053
0.5	0.3	1	0.5	0.299(0.029)	0.300(0.029)	0.299(0.029)	0.300(0.029)	0.500(0.061)	0.301(0.065)	0.050
0.5	0.3	1	1	0.300(0.029)	0.299(0.029)	0.299(0.029)	0.299(0.029)	0.500(0.060)	0.300(0.065)	0.050
0.5	0.3	1	2	0.299(0.029)	0.300(0.029)	0.300(0.028)	0.299(0.029)	0.499(0.060)	0.298(0.065)	0.049
0.5	0.3	1	4	0.300(0.029)	0.300(0.029)	0.300(0.029)	0.299(0.029)	0.500(0.059)	0.299(0.064)	0.048
0.5	0.3	1	10	0.299(0.029)	0.300(0.029)	0.300(0.029)	0.299(0.029)	0.499(0.060)	0.300(0.065)	0.049
Targeting optimal allocations in Theorem 1										
0	0	-	-	0.299(0.029)	0.299(0.028)	0.300(0.030)	0.299(0.028)	0.001(0.063)	0.001(0.062)	0.050
0.3	0.3	-	-	0.299(0.030)	0.299(0.029)	0.300(0.030)	0.299(0.028)	0.300(0.066)	0.300(0.065)	0.053
0.5	0.3	-	-	0.299(0.030)	0.299(0.029)	0.300(0.030)	0.299(0.028)	0.500(0.061)	0.300(0.064)	0.052

Table 3.4.2: Comparison of complete randomization and DBCD targeting the optimal allocation in Theorem 2 in terms of EWNF* and power

ρ_A	ρ_B	N	w_1 / w_2	1 / 0.1	1 / 0.25	1 / 0.5	1 / 1	1 / 2	1 / 4	1 / 10	
$p_{A1} = 0.20$ $p_{B1} = 0.35$ $p_{A2} = 0.30$ $p_{B2} = 0.45$											
0	0	178	CR	140	157	185	241	352	575	1244	
			RAR	139	155	183	238	348	568	1228	
			Power	0.803	0.801	0.803	0.806	0.797	0.806	0.802	(0.794**)
0.1	0.1	197	CR	155	174	204	266	390	636	1377	
			RAR	153	171	201	262	383	625	1349	
			Power	0.802	0.809	0.803	0.794	0.805	0.802	0.808	(0.798)
0.3	0.1	210	CR	165	185	218	284	415	678	1467	
			RAR	164	183	215	280	410	669	1444	
			Power	0.803	0.803	0.809	0.805	0.807	0.803	0.804	(0.800)
0.3	0.3	234	CR	184	206	243	316	463	756	1634	
			RAR	182	203	239	311	455	742	1606	
			Power	0.804	0.806	0.802	0.802	0.805	0.808	0.805	(0.803)
0.5	0.3	250	CR	197	221	260	338	494	807	1745	
			RAR	195	218	256	333	488	796	1722	
			Power	0.797	0.800	0.804	0.803	0.802	0.803	0.803	(0.799)
$p_{A1} = 0.25$ $p_{B1} = 0.25$ $p_{A2} = 0.45$ $p_{B2} = 0.30$											
0	0	395	CR	322	359	420	544	791	1285	2768	
			RAR	320	357	419	541	785	1273	2735	
			Power	0.806	0.801	0.807	0.803	0.799	0.808	0.805	(0.798)
0.1	0.1	395	CR	321	359	420	544	791	1285	2768	
			RAR	320	357	419	541	785	1273	2736	
			Power	0.805	0.801	0.802	0.799	0.803	0.806	0.802	(0.802)
0.3	0.1	380	CR	309	345	405	524	761	1237	2665	
			RAR	308	344	403	521	756	1226	2637	
			Power	0.811	0.807	0.797	0.805	0.803	0.808	0.808	(0.800)
0.3	0.3	360	CR	293	327	383	496	721	1171	2522	
			RAR	292	325	381	493	716	1160	2496	
			Power	0.806	0.806	0.799	0.806	0.803	0.804	0.805	(0.801)
0.5	0.3	334	CR	272	303	356	460	669	1087	2340	
			RAR	271	302	354	458	666	1080	2321	
			Power	0.807	0.806	0.803	0.796	0.807	0.796	0.805	(0.798)

*For each combination of ρ_A and ρ_B, the first line is EWNF of CR procedure, and the second line is EWNF of RAR procedure.

**For each combination of ρ_A and ρ_B, the third line is the power of DBCD as well as the power of CR that is reported in parentheses for comparison.

3.5 Re-designing the trial of perioperative total parenteral nutrition (TPN) treatment using the RAR design

We also redesign the trial mentioned in Section 2.7 with RAR targeting the optimal allocation in obtained from Thereom 2. In Table 3.5.1, we report the results of DBCD aiming to minimize the EWNF while maintaining the power, and we compare it with CR. In each scenario, we consider the correlations between the endpoints to range from 0 to 0.5, and we adjust the total sample size to achieve 80% power for CR. Table 2.7.1 shows that the RAR design can either achieve higher power of detecting the treatment effects under different levels of correlation between the endpoints, or reduce the number of failures while holding the power to the same level as that for CR.

Table 3.5.1: Comparison of complete randomization and DBCD targeting the optimal allocation in Theorem 2 for redesigned trial in terms of EWNF* and power

ρ_A	ρ_B	N	w_1 / w_2	1 / 0.1	1 / 0.25	1 / 0.5	1 / 1	1 / 2	1 / 4	1 / 10	
0	0	240	CR	207	229	266	342	492	792	1692	
			RAR	205	227	264	338	485	779	1665	
			Power	0.809	0.808	0.807	0.806	0.799	0.804	0.804	(0.802**)
0.1	0.1	260	CR	225	248	289	370	533	858	1834	
			RAR	222	246	286	366	525	845	1802	
			Power	0.800	0.803	0.798	0.802	0.807	0.799	0.805	(0.799)
0.2	0.2	285	CR	246	272	317	406	584	941	2010	
			RAR	243	269	313	401	575	925	1975	
			Power	0.802	0.808	0.810	0.810	0.804	0.802	0.806	(0.801)
0.3	0.1	240	CR	242	267	311	399	574	924	1975	
			RAR	239	264	307	393	564	907	1935	
			Power	0.801	0.797	0.800	0.797	0.796	0.798	0.798	(0.796)
0.3	0.3	240	CR	259	287	334	427	615	991	2118	
			RAR	256	283	329	421	605	973	2076	
			Power	0.798	0.800	0.800	0.797	0.801	0.801	0.798	(0.799)
0.5	0.3	240	CR	290	320	372	477	687	1106	2364	
			RAR	286	316	367	470	674	1083	2308	
			Power	0.808	0.811	0.810	0.809	0.806	0.811	0.804	(0.795)

*For each combination of ρ_A and ρ_B, the first line is EWNF of CR procedure, and the second line is EWNF of DBCD.

**For each combination of ρ_A and ρ_B, the third line is the power of DBCD as well as the power of CR that is reported in parentheses for comparison.

3.6 Discussion

The trade-off between the power and the EWNF has long been discussed. We show that there exist procedures where the power is not compromised when more people are assigned to better treatment. We also see that the extent of improvement of RAR over traditional clinical trial designs depends on lots of factors such as the objectives and parameter settings. In some cases, the improvement is not quite significant. But it is also worth noting that any improvement is desirable for certain deadly diseases such as Ebola. One of the major contribution of the paper is to derive the analytic optimal allocation for clinical trials with two possibly correlated endpoints. It is not only useful to understand the procedure but also provides foundation for future theoretical study and comparison.

The promotion of RAR procedures in real clinical trials require extensive research on its application in diverse practical scenarios. First, clinical trials with more than two endpoints could be explored using a similar framework. However, more endpoints may require more parameters and a much more complicated covariance matrix. Second, other response types or a mixed type of endpoints such as one continuous and one binary endpoint could also be investigated. Third, when there are primary hypotheses and secondary hypotheses and the secondary hypotheses will be tested only when the primary hypotheses are rejected, the graphical approach to sequential testing (Bretz et al. (2009)) is a useful tool for fixed design. When RAR procedures are used in this case, the graphical approach is potentially helpful to understand and define the objective of the problem and write out the objective function in the optimization problem. Fourth, discussed different types of powers, and innovative approaches such as using a latent variable were proposed. When RAR procedures are used in the above scenarios, we need to write out different objective or utility functions in the optimality problems. It may lead to difficulty in deriving the closed form of the optimal allocation proportions. However, we can still obtain the numerical solutions for the target allocations and implement them with appropriate randomization procedures. Since the RAR design targets the theoretically derived optimal allocation proportions, it usually performs better than the balanced complete randomization in terms of the targeted objective, unless the balanced allocation proportion is the theoretically optimal one. Moreover, we can also use RAR procedures to achieve some desirable features such as assigning more patients to the better treatment even if they are not from formal optimality problems. The control of type I error rate could be addressed using the theoretical results from Hu and Zhang (2004) and the innovative methods proposed for the scenario of interest such as Bretz et al. (2009) and Senn and Bretz (2007). In recent years the use of Bayesian adaptive design has become popular in clinical development. They incorporate prior information naturally as

prior distribution and multiple interim analysis will not affect the final bayesian descision. (Yin (2012)). However Bayesian adaptive designs are usually are situation-specific. They rely heavily on model assumptions and simulation thus and cannot be generalized to all trials. Finally, responses are assumed to be available immediately in this paper. In certain real clinical trials where the responses become available after most patients have been recruited and randomized, the RAR procedures are not appropriate in general. However, there is no logical difficulty applying RAR procedures in trials with moderate delayed response, given that some responses become available during the randomization period. We leave these topics for future research.

Chapter 4

Testing Hypotheses of Covariate-Adaptive Randomized Clinical Trials with time to event outcomes

4.1 Introduction

With the development of Bioinformatics, it is well accepted that many covariates (biomarkers) are associated with certain diseases, and a variety of advanced concepts and methodologies including personalized medicine were proposed to address this issue. However, in clinical trials involving important covariates, the first and most common concern of clinical trial practitioners and physicians is still the balancing of these covariates for a simple treatment comparison. As a result, the covariate adaptive randomization (CAR) design including the stratified permuted block randomization (Taves (1974)) and Pocock and Simon's design (Pocock and Simon (1975)), which sequentially assigns the next patient to different treatment arms based on previous treatment assignments and covariates in order to balance the patients prognostic factors in each treatment arm, is definitely the common practice in modern clinical trials. This will avoid the inaccuracy introduced into the estimation of treatment effects due to the poor balance in patients' characteristics

(FDA (2010)). Balance in prognostic factors across treatments is also desirable for trials which have small sample size, involve interim analysis, or require subgroup analysis, etc. (Toorawa et al. (2009)). In addition, after identifying subgroups to develop a personalized medicine, covariate adaptive randomized clinical trial is often the next step to confirm the subgroups (Hu et al. (2015)). An overview of CAR designs can be found in Rosenberger and Sverdlov (2008) and Hu et al. (2015).

Other CAR designs with different properties can be found in Nordle and Brantmark (1977), Wei and Durham (1978), Signorini et al. (1993), Heritier et al. (2005), Russell et al. (2011), Hu et al. (2012), Lebowitsch et al. (2012), etc. However, all these papers focused on the theoretical and numerical properties of the CAR designs such as the extent of imbalance. Very little theoretical work has been done about the inference following CAR designs due to its complicated correlation structure of the within-stratum imbalances and the discreteness of the allocation probability function. Practitioners would typically use the same theory and statistical methods as if complete randomization was applied, which would lead to conservative of the Type I error rate if not all the randomization covariates were included in the data analysis. Extensive numerical studies have been conducted to confirm this problem (see, e.g. Birkett (1985), Forsythe (1987), Aickin (2002), Weir and Lees (2003), Hagino et al. (2004)). Until recently, Shao et al. (2010) and Ma et al. (2015) offered theoretical explanation for linear regression models. However, survival responses are the most commonly used endpoint in real trials. In this paper, we focus on the the statistical inference and control of type I error rate for clinical trials with survival analysis following CAR designs.

A lot of clinical trials employed stratified permuted blocks method in randomizing the patients (see. e.g, Van Der Graaf et al. (2012),Vestbo et al. (2016),van Gijn et al. (2011), and Rittmeyer et al. (2017)). In most of these clinical trials the common practice is to use Kaplan-Meier approach to estimate the median overall survival and to estimate

the hazard ratio with a stratified Cox regression analysis where stratification factors were the same used for randomization. The accelerated failure time model has been used for traditional survival data under independent right censoring as a useful alternative to the proportional hazards model. The inference procedures and their asymptotic properties have recently been intensively studied (Buckley and James (1979), Koul et al. (1981), Tsiatis (1990), Miller and Halpern (1982), Ritov (1990), Lai and Ying (1991), Ying (1993), Lin and Ying (1995), Jin et al. (2003), Leon et al. (2009)). Several approaches can be employed to estimate the parameters in the AFT model including Buckley-James estimator, rank-based estimator and the weighted least squares approach estimator (Stute (1993, 1996)), which is equivalent to the inverse probability weighting (IPW) estimator. Orbe et al. (2002) showed Stute's method can be applied when Cox assumptions also hold and be more precise in a simulated gastric cancer study. However, properties of these estimating methods on data from CAR designs have not explicitly investigated. Bandyopadhyay et al. (2010) discussed several exact and limiting properties of a covariate-adjusted adaptive design for two-stage clinical trials based on maximum likelihood method of AFT model. But they did not discuss the scenario when only a part of the randomization covariates are used in data analysis. In practice, it may not be possible to include all covariates in the working model and complex model will also increase the chance of misspecifying the model. In the Carotid Revascularization Endarterectomy versus Stenting Trial (CREST) described in Lancet Neurology (Lal et al. (2012)), the stratified permuted block design was used to balance allocation over center and symptomatic status. A parametric AFT model was fitted as a sensitivity analysis to support the lack of evidence of a treatment difference in restenosis rates. It is our motivation to investigate the statistical inference, especially the control of type I error rate, of a clinical trial with AFT model and CAR designs.

In this chapter, we answer the following two questions: (1) how the CAR design

affects the type I error rate in survival analysis using AFT models? (2) Can we perform formal statistical tests and fix the possible problems of type I error rate based on the derived theoretical results?

Extensive simulations are performed for the following three scenarios: (1) all the covariates used in CAR procedures are included in the data analysis, (2) only partial covariates are used in data analysis, (3) and no covariates other than treatment assignments are used in data analysis. When a simple treatment comparison is the major aim in a clinical trial, the latter two scenarios are very common for simplifying modeling technique and easing explanation of results. In addition, correct model specifications are often unknown in practice. We find that, under all three scenarios the parameter estimators are consistent, though CAR design is a complicated procedure and all the responses and treatment assignments are correlated with each other in a complex manner.

We found that under scenario (1) hypothesis testing using unadjusted test statistics is valid in terms of Type I error. Under scenario (2) and (3) conservative Type I error rate will be obtained if we make no adjustments to the test statistics. With less covariates incorporated in data analysis the conservation will become more serous. With our proposed adjustments to the test statistics, valid Type I error can be obtained under different scenarios. The test under CAR is more powerful than complete randomization when the treatment effects is relatively large and less powerful otherwise.

The remainder of this chapter is organized as follows. We introduce the framework and major theorems in Section 4.2 and Section 4.3. Numerical results are given in Section 4.4, and concluding remarks are offered in Section 4.5.

4.2 Framework

Consider a randomized clinical trial, in which n subjects are sequentially assigned one of two treatments using CAR designs. Let I_i, $i = 1, 2, \ldots, n$, indicates the assignment of the ith patient, i.e., $I_i = 1$ for treatment 1 and $I_i = 0$ for treatment 2. Let T be the survival time and C be the censoring time with the survival function $G(\cdot)$. Denote the covariates of interest by $\boldsymbol{X}_i = (X_{i,1}, X_{i,2} \ldots X_{i,p})^T$ and $\boldsymbol{Z}_i = (Z_{i,1}, Z_{i,2} \ldots Z_{i,q})^T$, and assume that the CAR designs are applied with respect to both $\boldsymbol{X}_i$s and $\boldsymbol{Z}_i$s, but only $\boldsymbol{X}_i$s are used in data analysis. In this paper, every covariate is assumed to be a scaler for simplicity, but all the results can be extended to the multi-dimensional covariate vectors. The observed data are represented by $(Y_i, \delta_i, \boldsymbol{X}_i, \boldsymbol{Z}_i), i = 1, 2, ..., n,$ where $Y_i = min(T_i, C_i), \delta_i = I(T_i \leq C_i)$, and $I(\cdot)$ is the indicator function here and in the sequel.

Recall that the AFT model relates the logarithms of the failure time to the covariates of interest through a linear form (Kalbfleisch and Prentice (2011), Cox and Oakes (1984)). Assume the ith subject's response follows the following AFT model,

$$\log T_i = \mu_1 I_i + \mu_2(1 - I_i) + \beta_1 X_{i,1} + \cdots + \beta_p X_{i,p} + \gamma_1 Z_{i,1} + \cdots + \gamma_q Z_{i,q} + \sigma\varepsilon_i \qquad (4.1)$$

where μ_1 and μ_2 are parameters measuring the main effects of treatments 1 and 2, respectively, $(\beta_1, \ldots, \beta_p)$ and $(\gamma_1, \ldots, \gamma_q)$ are unknown parameters, σ is a scale parameter and ε_i are the random disturbance terms, usually assumed to be independent and identically distributed with some density function $f(\varepsilon)$. Here we assume ε_i are independent and identically distributed random errors with mean zero and variance σ_ε^2, and $X_{i,k}$ and $Z_{i,j}$, $k = 1, \ldots, p$, $j = 1, \ldots, q$ are independent and identically distributed as X_k and Z_j, respectively. Assume all covariates are independent of each other and the random errors, and without loss of generality assume $EX_k = 0$, $k = 1, \ldots, p$, and $EZ_j = 0$, $j = 1, \ldots, q$.

Conditional on the covariates, the censoring time and the survival time are assumed to be independent.

Let $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)^T$, $\mathbf{T} = (T_1, T_2, \ldots, T_n)^T$, $\boldsymbol{\beta} = (\mu_1, \mu_2, \beta_1, \ldots, \beta_p)^T$ $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_q)^T$,

$$\boldsymbol{X} = \begin{bmatrix} I_1 & 1-I_1 & X_{1,1} & \ldots & X_{1,p} \\ I_2 & 1-I_2 & X_{2,1} & \ldots & X_{2,p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ I_n & 1-I_n & X_{n,1} & \ldots & X_{n,p} \end{bmatrix} \text{ and } \boldsymbol{Z} = \begin{bmatrix} Z_{1,1} & \ldots & Z_{1,q} \\ \vdots & \ddots & \vdots \\ Z_{n,1} & \ldots & Z_{n,q} \end{bmatrix}$$

Then model (4.1) can be written as

$$\log\mathbf{T} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\gamma} + \sigma\boldsymbol{\varepsilon},$$

and the working AFT model is

$$E[\log T_i] = \mu_1 I_i + \mu_2(1-I_i) + \beta_1 X_{i,1} + \cdots + \beta_p X_{i,p} \tag{4.2}$$

From the weighted least squares method (Stute (1993)), we have

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\ln\mathbf{T} \tag{4.3}$$

where the Kaplan–Meier weights $\mathbf{W}$ are

$$W_{in} = \hat{G}_n\left(\log T_{(i)}\right) - \hat{G}_n\left(\log T_{(i-1)}\right) = \frac{\delta_{[i]}}{N-i+1}\prod_{j=1}^{i-1}\left(\frac{N-j}{N-j+1}\right)^{\delta_{[j]}} \tag{4.4}$$

Let $T_{(i)}$ be the ith ordered value of the observed response variable, $\hat{G}_n$ is a Kaplan–Meier estimator of the distribution function for the variable T, and $\delta_{[i]}$ is the value

associated with $\log T_{(i)}$.

In this paper, we discuss clinical trials with treatment comparison, and consider the following hypothesis test:

$$H_0 \ : \mu_1 - \mu_2 = 0 \text{ versus } H_A \ : \mu_1 - \mu_2 \neq 0. \tag{4.5}$$

The test statistics for (4.5) is

$$\mathcal{T}(n) = \frac{\boldsymbol{L}\hat{\boldsymbol{\beta}}}{\{\hat{\text{Var}}(\boldsymbol{L}\hat{\boldsymbol{\beta}})\}^{1/2}} \tag{4.6}$$

where $\boldsymbol{L} = (1, -1, 0, .., 0)$.

In clinical trials, CAR designs are usually based on discrete covariates (Taves (2010)). If a continuous covariate is to be used in randomization, we need to discretize this variable. Let $C^* = \{k | X_k \text{ is continuous}, k = 1, \ldots, p\}$, $C = \{j | Z_j \text{ is continuous}, j = 1, \ldots, q\}$, $d_k^*(X_k)$ and $d_j(Z_j)$ be discrete functions,

$$\tilde{X}_k = \begin{cases} X_k & \text{if } k \notin C^* \\ d_k^*(X_k) & \text{if } k \in C^* \end{cases}$$

and

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

The CAR design will be applied with respect to $\tilde{X}_{i,k}$ and $\tilde{Z}_{i,j}$.

Suppose $\tilde{X}_k$ has s_k^* levels and $\tilde{Z}_j$ has s_j levels. Let $W_i = \Big(\tilde{X}_{i,1}, \ldots, \tilde{X}_{i,p}, \tilde{Z}_{i,1}, \ldots, \tilde{Z}_{i,q}\Big)$ represent the ith patient's covariate profile used in CAR designs. We use $(t_1, t_2, \ldots, t_p, r_1, r_2, \ldots, r_q)$ to denote the stratum formed by patients who have the same covariate levels $x_k^{t_k}$ for $\tilde{X}_k, k = 1, \ldots, p$ and $z_j^{r_j}$ for $\tilde{Z}_j, j = 1, \ldots, q$, use $(k; t_k)$ to denote

the margin formed by patients with level $x_k^{t_k}$ for covariate $\tilde{X}_{i,k}$, and $(j; r_j)$ to denote the margin formed by patients with level $z_j^{r_j}$ for covariate $\tilde{Z}_{i,j}$. Then denote

1. D_n be the difference between the numbers of patients in treatment group 1 and 2;
2. $D_n^*(k; t_k)$ and $D_n(j; r_j)$ be the differences between the numbers of patients in the two treatment groups on the margin $(k; t_k)$ and $(j; r_j)$, respectively;
3. $D_n(t_1, t_2, \ldots, t_p, r_1, r_2, \ldots, r_q)$ be the difference between the numbers of patients in the two treatment groups within the stratum $(t_1, t_2, \ldots, t_p, r_1, r_2, \ldots, r_q)$.

4.3 Method and Main results

When CAR designs are applied in the clinical trials, the major concerns are whether traditional tests such as t-test are still valid with well-controlled type I error rate, and whether the power is adversely affected, due to the dependence among responses, treatment assignments and covariates. The primary purpose of this section is to implement test statistics 4.6 with CAR designs to check the impact of CAR on inference and hypothesis testing process. A test is said to be (asymptotically) conservative, if the true Type I error is smaller than the significance level under the null hypothesis.

Finding 1 Suppose that a covariate adaptive design satisfies the following two conditions:

(A) the overall imbalance is bounded in probability, i.e., $D_n = O_P(1)$;

(B) the marginal imbalances for all covariates are bounded in probability, i.e., $D_n^*(k; t_k) = O_P(1)$ and $D_n(j; r_j) = O_P(1), k = 1, \ldots, p,\ j = 1, \ldots, q$.

Then, if the test statistic (4.6) is used to perform the hypothesis test (4.5):

(1) A valid type I error rate can be obtained if all the randomization covariates are included in the data analysis. Consistent estimation of the treatment effect can also be obtained with weighted least squared method.

(2) The type I error rate will be conservative if not all the randomization covariates are included in the data analysis. That is, for a given significance level α, there is a constant α_0 such that, when H_0 holds, $\lim_{n\to\infty} \mathrm{P}(|\mathcal{T}(n)| > Z_{1-\alpha/2}) \leq \alpha_0 < \alpha$. However, estimation of the treatment effect is also valid.

(3) The change of power will depend on the difference of treatment effects $(\mu_1 - \mu_2)$ if not all the randomization covariates are included in the data analysis.

Remark (1) Various CAR designs such as Pocock and Simon's design and the stratified permuted block design meet the mild conditions that the overall and marginal imbalances in patient numbers between the treatments are bounded in probability.

Simulation results show that, for any observed design matrix from a CAR procedure, when all randomization covariates are included to construct the test, the test statistics will asymptotically follow a $N(0,1)$ distribution hence acheive valid Type I error level, which validates all the previous clinical trials using this methods. On the other hand, excluding some covariates used in the randomization procedure from constructing the test, variance of the test statistics (4.6) is always less than 1, leading to a conservative Type I error rate, which explains the problem of many clinical trials using CAR designs but ignoring some randomization covariates such as clinic sites. This echoes Forsythe (1987)'s recommendation that "all variables used in minimization are also to be used as covariate" under simple linear model framework.

It is of importance to obtain a consistent estimator $\hat{\varepsilon}$ of the variance of $\boldsymbol{L}\hat{\boldsymbol{\beta}}$ in order to control the Type I error rate under scenario (2) of Finding 1. Bootstrap methods are proposed to directly obtain $\hat{\varepsilon}$.

4.3.1 Adjusting test statistics using Bootstrap

As we discussed above we may get a conservative Type I error if we do not incorporate all the covariates while this is not uncommon in practice. Shao et al. (2010) proposed a consistent variance estimator using the bootstrap method for linear models. It bootstrapped the original dataset and applied same randomization procedure to every bootstrap dataset to regenerate the treatment assignments. We applied similar method to our survival models to get the estimated variance of $\mathbf{L}\boldsymbol{\beta}$, we briefly discuss the bootstrap methods with to obtain a valid test. For simplicity of demonstration, we assume the response follows the following model with only two covariates,

$$\log T_i = \mu_1 I_i + \mu_2(1 - I_i) + \gamma_1 Z_{i,1} + \gamma_2 Z_{i,2} + \sigma\varepsilon_i,\ Y_i = \min(T_i, C_i) \tag{4.7}$$

We assume the CAR designs are applied with both covariates, but the final analysis is a simple treatment comparison without any covariates in the fitted AFT model. We generate bootstrap data $(Y_1^*, \delta_1^*, Z_{1,1}^*, Z_{1,2}^*), \ldots (Y_n^*, \delta_n^*, Z_{n,1}^*, Z_{n,2}^*)$ as a simple random sample with replacement from $(Y_1, \delta_1, Z_{1,1}, Z_{1,2}), \ldots, (Y_n, \delta_n,$
$Z_{n,1}, Z_{n,2})$. Apply the covariate-adaptive randomization procedure on the covariates of each bootstrap sample $(Z_{1,1}^*, Z_{1,2}^*), \ldots (Z_{n,1}^*, Z_{n,2}^*)$ to obtain the bootstrap analogues of treatment assignments $I_1^{*b}, \ldots, I_n^{*b}$. Then the bootstrap estimator of the variance of the (4.8) is $v_B = \mathrm{Var}_*(\hat{\theta}^*)$ where Var_* is the asymptotic variance of the bootstrap samples $\hat{\theta}^*$ and B is the number of independent bootstrap dataset we generate. In practice, v_B will be approximated by the sample variance of Var_*s from B bootstrap sample. The bootstrap datasets are simple random samples since we draw the samples independently with replacement. The consistency of these variance estimator will finally lead to a valid test with well controlled Type I error rate.

4.4 Simulation studies

In this section, we investigate finite sample performance of our proposed methods, and compare three randomization procedures, i.e., Pocock and Simon's design (PS), stratified permuted block design (SPB) and complete randomization (CR). First, we conduct simulation studies to empirically exam the performance of these randomization procedures. Three cases are considered:

Case I. Z_1 and Z_2 are discrete variables following Bernoulli(0.5)

Case II. Z_1 is discrete following Bernoulli(0.5) and Z_2 is continuous following Uniform$(0, 1)$. The CAR designs will be applied with respect to Z_1 and $\tilde{Z}_2$ as follows,

$$\tilde{Z}_2 = \begin{cases} 1 & \text{if } Z_2 < 0.5 \\ 0 & \text{if } Z_2 \geq 0.5, \end{cases} \tag{4.8}$$

Case III. Z_1 and Z_2 are from multivariate normal distribution $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\tilde{Z}_1, \tilde{Z}_2$ with similar definition as in 4.13 are used in CAR randomization procedures, where

$$\boldsymbol{\mu} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \boldsymbol{\Sigma} = \begin{pmatrix} 0.25 & 0.16 \\ 0.16 & 0.25 \end{pmatrix} \tag{4.9}$$

Mean and standard deviation(Std) for strata difference $D_n(1,1), D_n(2,2)$, marginal difference $D_n(1;1), D_n(2;2)$ and overall difference D_n are reported in Table 4.4.1

Table 4.4.1: Mean(Std) of $D_n(\cdot)$ of several methods under different sample sizes. Simulations based on 5000 runs

Sample size		$\mathbf{D_n(1,1)}$	$\mathbf{D_n(2,2)}$	$\mathbf{D_n(1;1)}$	$\mathbf{D_n(2;2)}$	$\mathbf{D_n}$
PS(Case I)	500	0.04(5.81)	0.04(5.82)	0.04(3.43)	0.04(3.4)	0.04(2.06)
	250	-0.01(4.34)	0.04(4.37)	-0.01(3.34)	0.04(3.39)	0.01(2.02)
SPB(Case I)	500	0.02(0.91)	-0.01(0.92)	0.02(1.27)	0.00(1.28)	0.02(1.79)
	250	-0.01(0.91)	-0.01(0.92)	-0.01(1.28)	-0.02(1.28)	-0.01(1.84)
PS(Case II)	500	0.09(5.77)	0.07(5.82)	0.01(3.45)	-0.01(3.41)	0.00(2.09)
	250	0.04(4.34)	0.04(4.3)	0.02(3.36)	0.02(3.34)	0.03(2.02)
SPB(Case II)	500	-0.03(0.89)	0.00(0.91)	0.01(1.29)	0.04(1.28)	0.00(1.80)
	250	-0.03(0.90)	0.01(0.92)	-0.02(1.28)	0.02(1.30)	0.00(1.81)
PS(Case III)	500	0.11(5.63)	0.13(5.66)	-0.02(1.76)	0.00(1.77)	-0.02(2.07)
	250	-0.07(4.07)	-0.05(4.09)	-0.01(1.74)	0.02(1.76)	-0.01(2.04)
SPB(Case III)	500	-0.07(5.70)	-0.06(5.67)	-0.02(1.73)	-0.01(1.8)	0.01(2.08)
	250	-0.08(4.14)	-0.06(4.14)	0.01(1.72)	0.04(1.74)	0.03(1.99)

The mean difference between the numbers of patients in two treatment groups, marginal difference and the difference within stratum all go to 0 which indicates good balance results from Pocock and Simon's design and stratified permuted block design.

Second, we study the effect of CAR designs on statistical inference in a clinical trial with survival analysis. We assume correct model 4.1 is specified with survival time T_i follows the following Accelerated Failure Time model,

$$\log(T_i) = \mu_1(1 - I_i) + \mu_2 I_i + \beta_1 Z_{i,1} + \beta_2 Z_{i,2} + \sigma\varepsilon_i, \tag{4.10}$$

where $\beta_1 = \beta_2 = 0.5, \sigma = 1$ and ε_i follows a normal distribution $N(0, 0.25)$. We also set $\mu_1 = \mu_2 = 0.5$ to study the Type I error rate. And we choose $c = 13$ and 7 resulting in censoring probability of approximately 15% and 30% for case 3.

For simplicity, we use only two covariates for numerical study and we do not distinguish the notation of $\boldsymbol{X}$ and $\boldsymbol{Z}$. The following cases are considered in the simulation:

Case 1. Z_1 follows Bernoulli(p_1) and Z_2 follows Bernoulli(p_1)

Case 2. Z_1 follows normal distribution $N(0, 0.25)$ and Z_2 follows normal distribution $N(0, 0.25)$.

Case 3. Z_1 follows Bernoulli(p_1) and Z_2 follows Uniform distribution $(0, 1)$.

In case 2 and case 3 the CAR designs are applied with discretized variable

$$\tilde{Z} = \begin{cases} 1 & \text{if } Z < p_2 \\ 0 & \text{if } Z \geq p_2, \end{cases} \tag{4.11}$$

The biased coin probability 0.75 and equal weights are used for Pocock and Simon's design, and the block size 4 is used for stratified permuted block design. The significance level is $\alpha = 0.05$. The following four models will be fit to analyze the data: (1) no covariates are included in the AFT model (AFT), (2) a single covariate Z_1 is included ($AFT(Z_1)$), (3) a single covariate Z_2 is included ($AFT(Z_2)$), (4) both covariates Z_1 and Z_2 are included in the AFT model ($AFT(Z_1, Z_2)$).

The simulation results are given in Tables 4.4.2, 4.4.3 for moderate sample size $n = 500$ and relatively small sample size $n = 250$, respectively. In this section, we let $p_1 = 0.8$ and $p_2 = 0.4$. Parameter estimates, type I error rate, and empirical standard deviations and estimated standard deviations of estimated treatment effects are reported. The Type I error rate is close to 5% with the full model $AFT(Z_1, Z_2)$ using CAR procedures and with all the four models using complete randomization, as predicted by our theorem. For both CAR designs here, the type I error rate is conservative with the AFT, $AFT(Z_1)$ and $AFT(Z_2)$ models where not all the randomization covariates are used in the data analysis. Moreover, the AFT model without any covariates returns the most conservative

Type I error rate.This is due to the variance of estimated treatment effects is inflated by CAR designs. It is also illustrated by the empirical standard deviations and estimated standard deviations in Tables 4.4.2, 4.4.3.

In Tables 4.4.4 and 4.4.5 we show the results with different right-censoring rates. Among these three tests,The inference results of the parameters using traditional AFT model are consistent under covariate-adaptive designs. When the right-censoring percentage is high (30%), the estimator of μ_1 and μ_2 can be biased due to the instability of weight function in the denominator. This phenomenon remains when the total sample size increases from 250 to 500. With larger sample size and less censoring we can get better estimation of treatment effect.

Third, we study the effect of CAR designs on statistical inference in a clinical trial with survival analysis when model is misspecified. We consider the following cases.

Case 4. The true model is $\log(T_i) = \mu_1(1 - I_i) + \mu_2 I_i + \beta_1 Z_{i,1} + \beta_2 Z_{i,2}^2 + \sigma\varepsilon_i$. where $\sigma = 1$ and ε_i follows a normal distribution $N(0, 0.25)$. The working model is $\log(T_i) = \mu_1(1 - I_i) + \mu_2 I_i + \beta_1 Z_{i,1} + \beta_2 Z_{i,2} + \sigma\varepsilon_i$ without recognizing that the effect of Z_2 is quadratic.

Case 5. When the distribution of error term follows the Gumbel distribution and the scale factor $\alpha = 1$ then the AFT model is the exponential distribution. In this case the expectation of ε_i, $E(\varepsilon_i) > 0$.

Case 6. When Z_1 and Z_2 are correlated with multivariate normal distribution $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where

$$\boldsymbol{\mu} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \boldsymbol{\Sigma} = \begin{pmatrix} 0.25 & 0.16 \\ 0.16 & 0.25 \end{pmatrix} \tag{4.12}$$

Case 7. Error terms are autocorrelated that is

$$\begin{aligned} \log(T_i) &= \mu_1(1-I_i) + \mu_2 I_i + \beta_1 Z_{i,1} + \beta_2 Z_{i,2} + \varepsilon_i, \\ \varepsilon_i &= \rho\varepsilon_{i-1} + u_i, -1 < \rho < 1, u_i \sim N(0, \sigma^2) \end{aligned} \tag{4.13}$$

where $\sigma^2 = 0.25$ is used.

Remark (2) In clinical trials since patients are often enrolled sequentially, it's very common that the patients arrive in batches. Patients with close enrollment time maybe correlated. That's why we consider the autocorrelated residuals.
The results for case 4-6 are shown in Tables 4.4.7 - 4.4.10. Table 4.4.11 reports the Type I error after bootstrap method is applied to adjust the variance estimation and test statistic. Several conclusions can be obtained as follows.

(1) When the effect of Z_2 is quadratic, the wrong estimation of treatment effect may be generated. When all covariates used in CAR are included in the working model Type I errors are still valid. Otherwise the test will be conservative.

(2) If the true model is Weibull or exponential the weighted least squared estimator of treatment effect may not be consistent. The underlying reason is that the assumption of the estimator proposed by Stute is that $\mathrm{E}(\varepsilon|\mathrm{X}) = 0$. While for Weibull distribution or exponential distribution the error terms have positive expectation. In terms of Type I error we have similar observation as in (1).

(3) When to covariates are correlated, we can still get valid treatment effect estimation. Valid Type I error can also be achieved by including all covariates that are used in CAR.

(4) When the error terms are slightly autocorrelated, as long as the expectation of ε_i, $E\varepsilon_i = 0$n the estimation of treatment effect is consistent. Valid Type I error

can also be achieved by including all covariates that are used in CAR.

(5) When bootstrap method is applied tp adjust the variance estimation, valid Type I error can still be achieved.

In this paper, we propose to use Bootstrap methods to estimate the variance of the estimated treatment effects as described in Section 4.3.1, and the results are given in Table 4.4.6. Table 4.4.6 demonstrates the success of our proposed methods. The Type I error rate is close to the nominal level 5% under both Pocock and Simon's design and the stratified permuted block design. The proposed approaches solve the problem that the variance of treatment effects are overestimated for the empirical ones. These results are of special importance since they offer guidance on future trials.

We also compare power for different methods using a smaller sample size, $n = 100$, and setting $(p_1, p_2) = (0.5, 0.5)$. Results for different values of treatment effects $(\mu_1 - \mu_2)$ are displayed in Figure 1. We only show the results for Pocock and Simon's design. Figure 1 shows that the AFT model returns the lowest power, the $AFT(Z_1, Z_2)$ model has the biggest power, and the $AFT(Z_1)$ and $AFT(Z_2)$ model are in the middle. Compared to complete randomization, the Pocock and Simon's design has smaller power when $|\mu_1 - \mu_2|$ is relatively small, but has larger power when $|\mu_1 - \mu_2|$ becomes larger.

Figure 4.4.1: Simulated power for Pocock and Simon's marginal procedure (PS) and complete randomization (CR). Simulation based on 10000 runs and number of patients $n = 100$.

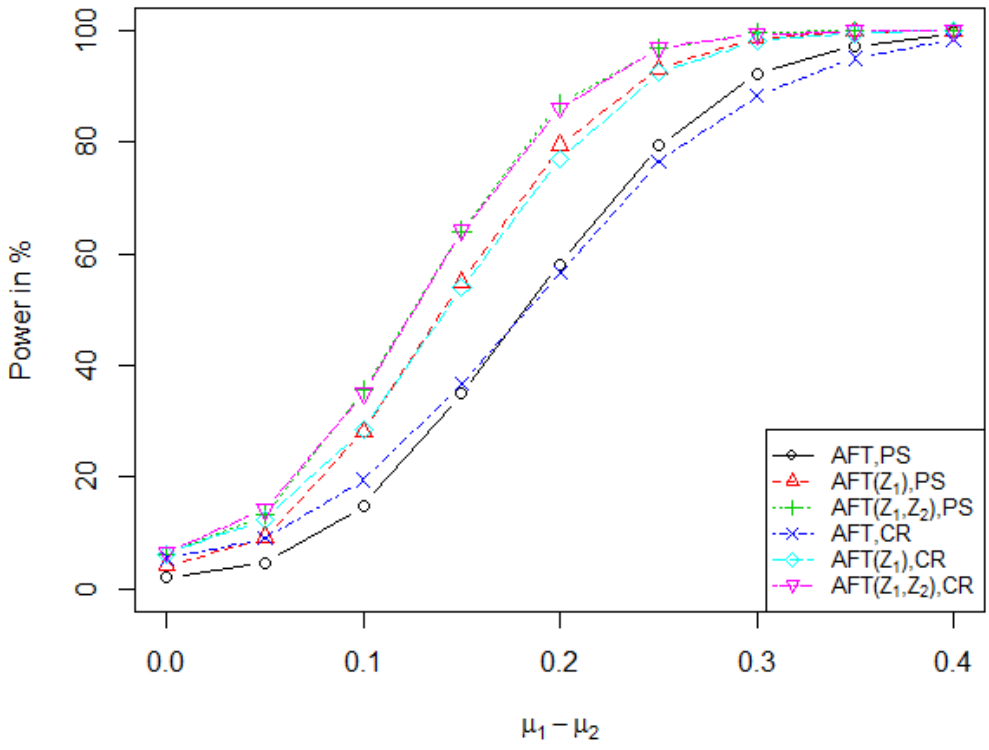

Table 4.4.2: Estimated coefficients and simulated Type I error in % when Z_1, Z_2 are discrete and independent. Simulations based on 5000 runs

Model	Method	μ_1	μ_2	β_1	β_2	Type I error	Emp. Std	Est. Std
				n=500, percentage of censoring 15%				
AFT	PS	.500	.499	-	-	.68	.032	.044
	SPB	.499	.499	-	-	.62	.031	.044
	CR	.500	.499	-	-	5.06	.044	.044
$AFT(Z_1)$	PS	.500	.499	.499	-	1.56	.028	.036
	SPB	.500	.500	.500	-	1.24	.028	.036
	CR	.500	.499	.499	-	5.24	.036	.036
$AFT(Z_2)$	PS	.499	.499	-	.499	1.42	.029	.036
	SPB	.500	.499	-	.499	1.34	.028	.036
	CR	.500	.499	-	.498	5.46	.036	.036
$AFT(Z_1, Z_2)$	PS	.500	.499	.500	.500	5.17	.026	.025
	SPB	.500	.500	.500	.499	5.12	.025	.025
	CR	.500	.500	.499	.499	5.28	.026	.025
				n=250, percentage of censoring 15%				
AFT	PS	.499	.499	-	-	.56	.045	.062
	SPB	.499	.499	-	-	.60	.044	.062
	CR	.498	.500	-	-	5.22	.063	.062
$AFT(Z_1)$	PS	.499	.499	.499	-	1.60	.040	.050
	SPB	.499	.500	.498	-	1.32	.040	.050
	CR	.498	.500	.498	-	5.32	.051	.050
$AFT(Z_2)$	PS	.499	.499	-	.499	1.38	.040	.050
	SPB	.498	.499	-	.500	1.30	.040	.050
	CR	.499	.500	-	.499	5.36	.051	.050
$AFT(Z_1, Z_2)$	PS	.500	.500	.499	.499	5.14	.036	.035
	SPB	.499	.500	.499	.500	5.22	.035	.035
	CR	.499	.499	.499	.500	5.00	.035	.035

Table 4.4.3: Estimated coefficients and simulated Type I error in % when Z_1, Z_2 are from normal distribution. Simulations based on 5000 runs

Model	Method	μ_1	μ_2	β_1	β_2	Type I error	Emp. Std	Est. Std
				n=500, percentage of censoring 15%				
AFT	PS	.499	.499	-	-	1.74	.040	.047
	SPB	.499	.500	-	-	2.42	.040	.047
	CR	.499	.499	-	-	5.10	.047	.047
$AFT(Z_1)$	PS	.499	.499	.498	-	2.66	.034	.039
	SPB	.499	.500	.498	-	2.64	.035	.039
	CR	.499	.499	.498	-	5.20	.039	.039
$AFT(Z_2)$	PS	.499	.500	-	.499	2.62	.035	.039
	SPB	.499	.500	-	.498	2.88	.034	.039
	CR	.499	.500	-	.498	5.20	.039	.039
$AFT(Z_1, Z_2)$	PS	.499	.500	.498	.499	5.12	.029	.029
	SPB	.500	.500	.499	.499	4.98	.029	.029
	CR	.500	.500	.499	.499	5.40	.029	.029
				n=250, percentage of censoring 15%				
AFT	PS	.499	.498	-	-	2.24	.057	.066
	SPB	.498	.498	-	-	2.00	.056	.066
	CR	.498	.498	-	-	5.22	.067	.066
$AFT(Z_1)$	PS	.499	.498	.496	-	2.58	.049	.054
	SPB	.498	.498	.497	-	2.94	.049	.054
	CR	.498	.498	.498	-	5.24	.056	.054
$AFT(Z_2)$	PS	.499	.498	-	.496	3.30	.050	.054
	SPB	.498	.499	-	.496	2.58	.048	.055
	CR	.499	.499	-	.496	5.36	.055	.054
$AFT(Z_1, Z_2)$	PS	.499	.499	.498	.498	5.18	.041	.040
	SPB	.498	.499	.499	.498	5.26	.041	.041
	CR	.499	.499	.499	.497	5.38	.041	.040

Table 4.4.4: Estimated coefficients and simulated Type I error in % when Z_1 is discrete and Z_2 is continuous. Simulations based on 5000 runs and Sample Size $n = 500$

Model	Method	μ_1	μ_2	β_1	β_2	Type I error	Emp. Std	Est. Std
				n=500, percentage of censoring 15%				
AFT	PS	.500	.500	-	-	1.62	.032	.044
	SPB	.499	.499	-	-	1.96	.031	.044
	CR	.500	.500	-	-	5.72	.045	.044
$AFT(Z_1)$	PS	.500	.499	.500	-	3.82	.030	.037
	SPB	.500	.500	.500	-	4.38	.030	.037
	CR	.501	.500	.500	-	5.48	.037	.037
$AFT(Z_2)$	PS	.500	.500	-	.501	1.88	.030	.037
	SPB	.499	.500	-	.501	2.20	.030	.037
	CR	.500	.500	-	.502	5.54	.037	.037
$AFT(Z_1, Z_2)$	PS	.500	.500	.500	.500	5.34	.028	.028
	SPB	.500	.500	.500	.500	5.82	.028	.028
	CR	.500	.499	.500	.500	5.02	.028	.028
				n=500, percentage of censoring 30%				
AFT	PS	.499	.499	-	-	1.34	.041	.051
	SPB	.499	.499	-	-	1.44	.041	.051
	CR	.500	.499	-	-	5.74	.052	.051
$AFT(Z_1)$	PS	.499	.499	.499	-	2.24	.036	.042
	SPB	.499	.500	.499	-	2.44	.036	.042
	CR	.500	.500	.499	-	5.58	.043	.042
$AFT(Z_2)$	PS	.499	.500	-	.500	2.26	.036	.042
	SPB	.499	.499	-	.499	2.50	.036	.042
	CR	.500	.499	-	.500	5.66	.043	.042
$AFT(Z_1, Z_2)$	PS	.500	.499	.500	.500	5.20	.031	.031
	SPB	.500	.500	.500	.500	4.96	.031	.031
	CR	.500	.500	.500	.500	5.52	.032	.031

Table 4.4.5: Estimated coefficients and simulated Type I error in % when Z_1 is discrete and Z_2 is continuous. Simulations based on 5000 runs and Sample Size $n = 250$

Model	Method	μ_1	μ_2	β_1	β_2	Type I error	Emp. Std	Est. Std
				n=250, percentage of censoring 15%				
AFT	PS	.500	.501	-	-	1.50	.044	.056
	SPB	.500	.501	-	-	1.16	.044	.056
	CR	.499	.499	-	-	5.64	.056	.056
$AFT(Z_1)$	PS	.500	.500	.500	-	3.62	.041	.044
	SPB	.500	.500	.500	-	3.98	.042	.044
	CR	.499	.499	.500	-	5.62	.044	.044
$AFT(Z_2)$	PS	.500	.501	-	.500	1.74	.042	.052
	SPB	.500	.500	-	.500	1.56	.042	.052
	CR	.499	.500	-	.499	5.48	.053	.052
$AFT(Z_1, Z_2)$	PS	.500	.500	.500	.499	5.02	.039	.039
	SPB	.500	.500	.500	.501	5.56	.040	.039
	CR	.499	.499	.501	.500	5.14	.040	.039
				n=250, percentage of censoring 30%				
AFT	PS	.499	.499	-	-	1.84	.053	.063
	SPB	.499	.499	-	-	2.06	.053	.063
	CR	.499	.499	-	-	5.78	.064	.063
$AFT(Z_1)$	PS	.500	.499	.499	-	4.20	.047	.049
	SPB	.499	.499	.499	-	4.56	.048	.049
	CR	.499	.499	.498	-	5.74	.049	.049
$AFT(Z_2)$	PS	.499	.499	-	.497	2.18	.050	.058
	SPB	.499	.499	-	.498	2.26	.050	.058
	CR	.499	.499	-	.499	5.82	.059	.058
$AFT(Z_1, Z_2)$	PS	.500	.499	.499	.498	5.50	.044	.043
	SPB	.499	.499	.499	.498	5.98	.044	.043
	CR	.499	.499	.498	.499	5.46	.044	.043

Table 4.4.6: Simulation Type I error in % when Z_1 is discrete and Z_2 is continuous from bootstrap. Simulation based on 5000 runs with 15% censoring, B=500

Model	Method	α	α_B	α	α_B
		n=500		n=250	
AFT	PS	1.38	4.96	1.50	4.98
	SPB	1.34	5.22	1.68	5.14
AFT(Z1)	PS	3.94	5.28	3.8	5.26
	SPB	3.46	5.58	3.78	5.42
AFT(Z2)	PS	1.60	5.12	1.76	5.29
	SPB	1.46	5.25	1.76	5.26

α-Type I error, α_B-Bootstrap Type I error

Table 4.4.7: Estimated coefficients and simulated Type I error in % with high order term in model. Simulations based on 5000 runs

Model	Method	μ_1	μ_2	β_1	β_2	Type I error	Emp. Std	Est. Std
		n=500, percentage of censoring 15%						
AFT	PS	.541	.542	-	-	1.68	.032	.039
	SPB	.541	.542	-	-	1.82	.032	.039
	CR	.541	.540	-	-	5.82	.055	.055
$AFT(Z_1)$	PS	.541	.542	.500	-	3.92	.027	.029
	SPB	.541	.542	.501	-	4.06	.071	.029
	CR	.541	.541	.500	-	5.24	.041	.041
$AFT(Z_2)$	PS	.541	.542	-	.001	1.74	.032	.039
	SPB	.541	.542	-	.003	1.86	.032	.039
	CR	.541	.540	-	-.002	5.14	.056	.054
$AFT(Z_1, Z_2)$	PS	.541	.542	.500	.000	4.86	.029	.029
	SPB	.541	.542	.501	.002	5.14	.029	.029
	CR	.541	.540	.499	-.001	5.32	.041	.041
		n=250, percentage of censoring 15%						
AFT	PS	.540	.542	-	-	1.64	.045	.055
	SPB	.541	.541	-	-	1.70	.045	.055
	CR	.541	.540	-	-	5.32	.055	.055
$AFT(Z_1)$	PS	.541	.542	.501	-	3.82	.036	.041
	SPB	.541	.541	.500	-	3.30	.036	.041
	CR	.541	.541	.500	-	5.24	.041	.041
$AFT(Z_2)$	PS	.540	.541	-	.000	1.66	.045	.055
	SPB	.541	.541	-	-.001	1.82	.046	.054
	CR	.541	.540	-	-.002	5.14	.056	.054
$AFT(Z_1, Z_2)$	PS	.540	.542	.501	.001	5.04	.041	.041
	SPB	.541	.541	.500	-.001	5.50	.042	.041
	CR	.541	.540	.499	-.001	5.32	.041	.041

Table 4.4.8: Estimated coefficients and simulated Type I error in % when underlying model follows Weibull distribution. Simulations based on 5000 runs

Model	Method	μ_1	μ_2	β_1	β_2	Type I error	Emp. Std	Est. Std
				n=500, percentage of censoring 15%				
AFT	PS	.638	.637	-	-	1.42	.042	.052
	SPB	.638	.637	-	-	1.76	.043	.052
	CR	.638	.638	-	-	5.58	.052	.052
$AFT(Z_1)$	PS	.639	.638	.492	-	2.20	.039	.044
	SPB	.638	.638	.493	-	2.76	.039	.044
	CR	.638	.639	.494	-	5.42	.045	.044
$AFT(Z_2)$	PS	.639	.638	-	.492	2.54	.039	.044
	SPB	.638	.638	-	.492	2.62	.039	.044
	CR	.638	.638	-	.492	5.26	.044	.044
$AFT(Z_1, Z_2)$	PS	.639	.639	.493	.493	4.98	.035	.035
	SPB	.639	.639	.493	.492	5.88	.036	.035
	CR	.639	.639	.494	.493	5.26	.036	.035
				n=250, percentage of censoring 15%				
AFT	PS	.635	.636	-	-	1.44	.066	.080
	SPB	.635	.636	-	-	1.58	.065	.080
	CR	.636	.636	-	-	5.50	.082	.080
$AFT(Z_1)$	PS	.637	.637	.491	-	2.76	.060	.067
	SPB	.636	.637	.491	-	2.70	.060	.067
	CR	.637	.636	.491	-	5.28	.069	.067
$AFT(Z_2)$	PS	.636	.636	-	.490	2.56	.061	.067
	SPB	.636	.637	-	.489	2.56	.060	.068
	CR	.637	.637	-	.490	5.60	.069	.067
$AFT(Z_1, Z_2)$	PS	.638	.637	.491	.490	5.76	.056	.053
	SPB	.637	.638	.492	.490	5.76	.055	.053
	CR	.638	.637	.493	.492	5.3	.055	.053

Table 4.4.9: Estimated coefficients and simulated Type I error in % when Z_1, Z_2 are from multivariate normal distribution. Simulations based on 5000 runs

Model	Method	μ_1	μ_2	β_1	β_2	Type I error	Emp. Std	Est. Std
				n=500, percentage of censoring 15%				
AFT	PS	.498	.497	-	-	1.52	.046	.057
	SPB	.497	.497	-	-	1.46	.046	.057
	CR	.497	.498	-	-	5.16	.057	.057
$AFT(Z_1)$	PS	.499	.499	.815	-	4.50	.034	.035
	SPB	.499	.499	.815	-	3.96	.033	.035
	CR	.499	.499	.815	-	5.04	.035	.035
$AFT(Z_2)$	PS	.499	.499	-	.815	3.90	.033	.035
	SPB	.499	.499	-	.815	3.76	.034	.035
	CR	.499	.499	-	.815	5.14	.036	.035
$AFT(Z_1, Z_2)$	PS	.499	.499	.498	.498	5.60	.029	.029
	SPB	.499	.499	.499	.497	4.90	.029	.029
	CR	.499	.499	.498	.499	5.04	.029	.029
				n=250, percentage of censoring 15%				
AFT	PS	.496	.495	-	-	1.22	.064	.079
	SPB	.495	.496	-	-	1.50	.064	.079
	CR	.495	.496	-	-	5.10	.079	.079
$AFT(Z_1)$	PS	.498	.499	.815	-	4.54	.048	.050
	SPB	.498	.497	.813	-	4.32	.048	.050
	CR	.498	.498	.814	-	5.18	.050	.050
$AFT(Z_2)$	PS	.498	.498	-	.814	4.50	.047	.050
	SPB	.498	.497	-	.814	4.30	.047	.050
	CR	.499	.498	-	.813	5.44	.050	.050
$AFT(Z_1, Z_2)$	PS	.499	.499	.497	.498	5.62	.042	.041
	SPB	.498	.498	.498	.497	5.28	.041	.041
	CR	.499	.498	.499	.496	5.44	.041	.041

Table 4.4.10: Estimated coefficients and simulated Type I error in % with correlated error terms. Simulations based on 5000 runs

Model	Method	μ_1	μ_2	β_1	β_2	Type I error	Emp. Std	Est. Std
				n=500, $\rho = 0.1$				
AFT	PS	.500	.499	-	-	.56	.031	.044
	SPB	.499	.499	-	-	.64	.032	.044
	CR	.500	.499	-	-	5.14	.044	.044
$AFT(Z_1)$	PS	.500	.499	.498	-	1.44	.028	.036
	SPB	.500	.500	.500	-	1.30	.028	.036
	CR	.500	.499	.499	-	5.52	.036	.036
$AFT(Z_2)$	PS	.500	.499	-	.498	1.36	.028	.036
	SPB	.500	.499	-	.499	1.36	.029	.036
	CR	.499	.499	-	.499	5.58	.036	.036
$AFT(Z_1, Z_2)$	PS	.500	.500	.499	.499	5.54	.026	.025
	SPB	.500	.500	.500	.499	5.52	.025	.025
	CR	.500	.499	.500	.500	5.68	.026	.025
				n=500, $\rho = 0.5$				
AFT	PS	.499	.499	-	-	.66	.034	.046
	SPB	.500	.499	-	-	.54	.033	.046
	CR	.500	.499	-	-	5.20	.047	.046
$AFT(Z_1)$	PS	.499	.499	.499	-	1.54	.031	.038
	SPB	.500	.500	.498	-	1.32	.030	.038
	CR	.500	.500	.500	-	5.64	.038	.038
$AFT(Z_2)$	PS	.499	.499	-	.499	1.54	.031	.038
	SPB	.500	.499	-	.498	1.50	.030	.038
	CR	.500	.499	-	.500	5.26	.038	.038
$AFT(Z_1, Z_2)$	PS	.499	.499	.499	.499	4.98	.028	.028
	SPB	.500	.500	.499	.499	5.24	.027	.028
	CR	.500	.500	.500	.500	5.16	.029	.028

Table 4.4.11: Simulation Type I error in % for case 4-7 from bootstrap. Simulation based on 5000 runs with n=500, 15% censoring, B=500

Model	Model	case 4	case 5	case 6	case 7
AFT	PS	5.41	5.12	5.22	4.91
	SPB	5.12	4.78	5.75	4.98
AFT(Z1)	PS	5.32	5.08	4.58	5.33
	SPB	5.68	4.95	5.23	4.91
AFT(Z2)	PS	5.48	5.56	4.88	5.45
	SPB	5.09	4.92	4.68	5.03

4.5 Discussion

CAR designs, especially the stratified permuted block randomization designs, are the most commonly used randomization in clinical trials, since balancing treatment allocation for influential covariates is of special importance for various reasons. Two questions about CAR designs have been demanding solutions for a long time. First, what is the asymptotic imbalance at difference levels such as within strata, on the margins, or overall imbalances? Until recently, Hu et al. (2012)) used complicated Markov technique to derive the order of the imbalances of all these levels for a family of CAR designs. Ma et al. (2015) proved the asymptotic imbalance for both Pocock and Simon's marginal procedures (Pocock and Simon (1975)) and the stratified permuted block designs.

A second question is will the type I error rate be conservative if not all the covariates used in CAR designs are included in the data analysis for models with survival type outcomes. Shao et al. (2010) and Ma et al. (2015) obtained solution of this question for linear regression model. But the answer for survival analysis is still unrevealed. In this chapter we conducted extensive simulation studies of hypothesis testing under AFT model for clinical trials with time to event data using CAR designs. We compared the change of type I error and power when correct models are specified and all covariates

are included in the model with situations when the working models are not the true underlying model. We also showed that consistent estimation can be achieved from the inverse weighted estimating equation. We explicitly explains the conservative of Type I error rate if only some of the randomization covariates are used in the model. Bootstrap adjustment is also proposed to control Type I error rates.

This chapter opens a door to comprehensively study clinical trials with CAR designs and survival analysis, and it demonstrated feasibility of AFT model with weighted least squares estimator. It shed some light on the future directions to provide theoretical proof of the findings:

First, with CAR applied, the consistency of Kaplan-meier estimator will need to be prove while in many clinical trials the crude log-rank tests are used to compare the treatments. Second modern clinical trials often require interim analysis, and the joint distribution of the sequential statistics is needed. The theoretical asymptotic distribution of the will offer marginal distribution of each of these sequential statistics. .

Chapter 5

Appendices

5.1 Appendix of Paper 1 and 2

Proof of Theorem 1

This proof involves lengthy derivations and calculations; we emphasize the major ideas and steps. To prove the main theorem we need the following lemma.

Lemma 1. The noncentrality parameter $\phi(n_A, n_B)$ in (1) is a concave function, and $\nabla\phi \geq 0$.

We omit the proof of the concaveness of ϕ because it is similar to the proof in Tymofyeyev *et al.* (2007); we show that $\nabla\phi \geq 0$.

Let $\Delta_1 = p_{A1} - p_{B1}$, $\Delta_2 = p_{A2} - p_{B2}$, $A_1 = p_{A1}q_{A1}$, $A_2 = p_{A2}q_{A2}$, $B_1 = p_{B1}q_{B1}$, and $B_2 = p_{B2}q_{B2}$. Then

$$\phi(\mathbf{n}) = \left\{\Delta_1^2\left(\frac{A_2}{n_A} + \frac{B_2}{n_B}\right) + \Delta_2^2\left(\frac{A_1}{n_A} + \frac{B_1}{n_B}\right) - 2\psi\Delta_1\Delta_2\right\} \Bigg/ \left\{\left(\frac{A_1}{n_A} + \frac{B_1}{n_B}\right)\left(\frac{A_2}{n_A} + \frac{B_2}{n_B}\right) - \psi^2\right\} = \frac{a(\mathbf{n})}{S(\mathbf{n})}$$

where $\psi = \rho_A\sqrt{A_1A_2}/n_A + \rho_B\sqrt{B_1B_2}/n_B$.

$$\frac{\partial\psi}{\partial n_K} = -\rho_K\frac{\sqrt{K_1K_2}}{n_K^2},\ K = A, B$$

$$\frac{\partial a(\mathbf{n})}{\partial n_K} = -\frac{\Delta_1^2K_2}{n_K^2} - \frac{\Delta_2^2K_1}{n_K^2} + 2\Delta_1\Delta_2\rho_K\frac{\sqrt{K_1K_2}}{n_K^2},\ K = A, B$$

$$\frac{\partial S(\mathbf{n})}{\partial n_K} = -\frac{K_1}{n_K^2}\left(\frac{A_2}{n_A} + \frac{B_2}{n_B}\right) - \frac{K_2}{n_K^2}\left(\frac{A_1}{n_A} + \frac{B_1}{n_B}\right) + 2\psi\rho_K\frac{\sqrt{K_1K_2}}{n_K^2}\ K = A, B$$

$$\begin{aligned}\frac{\partial\phi(\mathbf{n})}{\partial n_K} =& \frac{1}{S(\mathbf{n})^2}\left\{\frac{\partial a(\mathbf{n})}{\partial n_K}S(\mathbf{n}) - a(\mathbf{n})\frac{\partial S(\mathbf{n})}{\partial n_K}\right\} \\ =& \left[\left\{K_2\psi^2 + K_1\left(\frac{A_2}{n_A} + \frac{B_2}{n_B}\right)^2 - 2\sqrt{K_1K_2}\left(\frac{A_1}{n_A} + \frac{B_1}{n_B}\right)\psi\rho_K\right\}M^2\right. \\ & \left. + K_1K_2\Delta_2^2\left\{\psi^2 - \left(\frac{A_1}{n_A} + \frac{B_1}{n_B}\right)\left(\frac{A_2}{n_A} + \frac{B_2}{n_B}\right)\right\}^2(1-\rho_K^2)\right]/S(\mathbf{n})^2n_K^2\end{aligned}$$

where $M = \Delta_1 - K_2\psi\Delta_2\left(A_1/n_A + B_1/n_B\right) - K_1\psi\Delta_2\left(A_2/n_A + B_2/n_B\right) + \sqrt{K_1K_2}\psi^2\Delta_2 + \sqrt{K_1K_2}\Delta_2\left(A_1/n_A + B_1/n_B\right)\left(A_2/n_A + B_2/n_B\right)\rho_K,\ K = A, B.$

We have that

$$\begin{aligned}& K_2\psi^2 + K_1\left(\frac{A_2}{n_A} + \frac{B_2}{n_B}\right)^2 - 2\sqrt{K_1K_2}\left(\frac{A_1}{n_A} + \frac{B_1}{n_B}\right)\psi\rho_K \\ &= \left\{\sqrt{K_2}\psi - \sqrt{K_1}\left(\frac{A_2}{n_A} + \frac{B_2}{n_B}\right)\right\}^2 + 2\sqrt{K_1K_2}\left(\frac{A_1}{n_A} + \frac{B_1}{n_B}\right)\psi(1-\rho_K) \geq 0\end{aligned}$$

when $0 \leq \rho_K \leq 1$.

Thus, we have $\nabla\phi \geq 0$.

For this convex problem, the first-order Karush-Kuhn-Tucker (KKT) necessary conditions for optimality become sufficient conditions. We need to find the Lagrange multi-

pliers and the optimal allocation **n*** satisfying the following KKT conditions:

$$\nabla\phi(\mathbf{n}^*) - \lambda\mathbf{w} + \lambda_A(\mathbf{e}_1 - D\mathbf{1}) + \lambda_B(\mathbf{e}_2 - D\mathbf{1}) = 0 \tag{A.1}$$

$$\lambda(\mathbf{w}'\mathbf{n}^* - \mathbf{N}) = \mathbf{0} \tag{A.2}$$

$$\lambda_K(n_K^* - D(n_A^* + n_B^*)) = 0,\ K = A, B \tag{A.3}$$

$$\lambda \geq 0,\ \lambda_K \geq 0,\ K = A, B \tag{A.4}$$

where $\mathbf{e}_j$ is a vector of all 0's except for a one in the jth position, $j = 1, 2$.

In Theorem 1 we let $\mathbf{w} = (1, 1)'$ and write condition (A.1) in terms of $\eta_A^* = n_A^*/(n_A^* + n_B^*)$. Condition (A.2) can always be satisfied by changing the scaling of the η_A^*'s. Then we have

$$\lambda_K(\eta_K^* - D) = 0,\ K = A, B \tag{A.5}$$

$$-\nabla\phi(\boldsymbol{\eta}*) + \lambda\mathbf{1} + \lambda_A \begin{pmatrix} D-1 \\ D \end{pmatrix} + \lambda_B \begin{pmatrix} D \\ D-1 \end{pmatrix} = 0. \tag{A.6}$$

When $\eta_A^* \neq 0, \eta_B^* \neq 0$, we substitute ϕ into (A.2) to get:

$$
\begin{aligned}
&\Big[\eta_B^4 \Big\{-A_1A_2^2\Delta_1^2(\rho_A^2-1) - A_1^2A_2\Delta_2^2(\rho_A^2-1) + 2A_1A_2\sqrt{A_1A_2}\Delta_1\Delta_2\rho_A(\rho_A^2-1)\Big\} \\
&+ \eta_A\eta_B^3 \Big\{-2A_1A_2B_2\Delta_1^2(\rho_A^2-1) - 2A_1A_2B_1\Delta_2^2(\rho_A^2-1) + 4A_1A_2\sqrt{B_1B_2}\Delta_1\Delta_2 \\
&\times(\rho_A^2-1)\rho_B\Big\} + \eta_A^3\eta_B\left\{2A_2B_1B_2\Delta_1^2(\rho_B^2-1) + 2A_1B_1B_2\Delta_2^2(\rho_B^2-1)\right. \\
&\left.-4\sqrt{A_1A_2}B_1B_2\Delta_1\Delta_2\rho_A(\rho_B^2-1)\right\} \\
&+ \eta_A^4\Big\{B_1B_2^2\Delta_1^2(\rho_B^2-1) + B_1^2B_2\Delta_2^2(\rho_B^2-1) - 2B_1B_2\sqrt{B_1B_2}\Delta_1\Delta_2\rho_B(\rho_B^2-1)\Big\} \\
&+ \eta_A^2\eta_B^2\left\{-A_2^2B_1\Delta_1^2 + A_1B_2^2\Delta_1^2 + A_2B_1^2\Delta_2^2 - A_1^2B_2\Delta_2^2 - A_1A_2B_2\Delta_1^2\rho_A^2\right. \\
&-A_1A_2B_1\Delta_2^2\rho_A^2 - 2\sqrt{A_1A_2}\sqrt{B_1B_2}(-A_2+B_2)\Delta_1^2\rho_A\rho_B \\
&+2A_1\sqrt{A_1A_2}\sqrt{B_1B_2}\Delta_2^2\rho_A\rho_B - 2\sqrt{A_1A_2}B_1\sqrt{B_1B_2}\Delta_2^2\rho_A\rho_B \\
&-2A_1A_2\sqrt{B_1B_2}\Delta_1\Delta_2(1+\rho_A^2)\rho_B + A_2B_1B_2\Delta_1^2\rho_B^2 + A_1B_1B_2\Delta_2^2\rho_B^2 \\
&-2A_2B_1\Delta_1\Delta_2(-\sqrt{A_1A_2}\rho_A + \sqrt{B_1B_2}\rho_B) - 2A_1B_2\Delta_1\Delta_2(-\sqrt{A_1A_2}\rho_A \\
&\left.+\sqrt{B_1B_2}\rho_B) + 2\sqrt{A_1A_2}B_1B_2\Delta_1\Delta_2\rho_A(1+\rho_B^2)\right\}\Big] \\
&\times\Big[\eta_B\Big\{-A_1(B_2\eta_A + A_2\eta_B - A_2\eta_B\rho_A^2) + 2\sqrt{A_1A_2}\sqrt{B_1B_2}\eta_A\rho_A\rho_B\Big\} \\
&+ B_1\eta_A\left\{-A_2\eta_B + B_2\eta_A(\rho_B^2-1)\right\}\Big]^{-2} = 0. \qquad (*)
\end{aligned}
$$

Setting $r = \eta_A/\eta_B$, with a simple calculation (*) becomes a quartic function: $ar^4 + br^3 + cr^2 + dr + e = 0$, where a, b, c, d, and e are as given in Theorem 1. We solve this quartic function and choose the roots between 0 and 1 to get the solutions.

For Theorem 2, we let $\mathbf{w} = (w_1q_{A1} + w_2q_{A2}, w_1q_{B1} + w_2q_{B2})'$ and obtain the solutions similarly.

Chapter 6

References

Aickin, M., 2002. Beyond randomization. The Journal of Alternative & Complementary Medicine 8, 765–772.

Aickin, M., 2009. A simulation study of the validity and efficiency of design-adaptive allocation to two groups in the regression situation. The International Journal of Biostatistics 5.

Antognini, A.B., Giovagnoli, A., 2004. A new 'biased coin design'for the sequential allocation of two treatments. Journal of the Royal Statistical Society: Series C (Applied Statistics) 53, 651–664.

Atkinson, A.C., 1982. Optimum biased coin designs for sequential clinical trials with prognostic factors. Biometrika 69, 61–67.

Bandyopadhyay, U., , A., Bhattacharya, R., 2010. A covariate-adjusted adaptive design for two-stage clinical trials with survival data. Statistica Neerlandica 64, 202–226.

Bandyopadhyay, U., Biswas, A., 2001. Adaptive designs for normal responses with prognostic factors. Biometrika 88, 409–419.

Birkett, N.J., 1985. Adaptive allocation in randomized controlled trials. Controlled clinical trials 6, 146–155.

Biswas, A., Bhattacharya, R., 2009. Optimal response-adaptive designs for normal responses. Biometrical Journal: Journal of Mathematical Methods in Biosciences 51, 193–202.

Biswas, A., Bhattacharya, R., 2010. An optimal response-adaptive design with dual constraints. Statistics & probability letters 80, 177–185.

Biswas, A., Mandal, S., 2004. Optimal adaptive designs in phase iii clinical trials for continuous responses with covariates, in: moDa 7—Advances in model-oriented design and analysis. Springer, pp. 51–59.

Brem, H., Piantadosi, S., Burger, P.C., Walker, M., Selker, R., Vick, N.A., Black, K., Sisti, M., Brem, S., Mohr, G., et al., 1995. Placebo-controlled trial of safety and efficacy of intraoperative controlled delivery by biodegradable polymers of chemotherapy for recurrent gliomas. The Lancet 345, 1008–1012.

Bretz, F., Maurer, W., Brannath, W., Posch, M., 2009. A graphical approach to sequentially rejective multiple test procedures. Statistics in medicine 28, 586–604.

Buckley, J., James, I., 1979. Linear regression with censored data. Biometrika 66, 429–436.

Cox, D.R., Oakes, D., 1984. Analysis of survival data. volume 21. CRC Press.

Dmitrienko, A., Tamhane, A.C., Bretz, F., 2009. Multiple testing problems in pharmaceutical statistics. CRC Press.

Efron, B., 1971. Forcing a sequential experiment to be balanced. Biometrika 58, 403–417.

Eisele, J.R., 1994. The doubly adaptive biased coin design for sequential clinical trials. Journal of Statistical Planning and Inference 38, 249–261.

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

FDA, 2010. Guidance for industry: Adaptive design clinical trials for drugs and biologics. Draft guidance .

FDA, CDER, CBER, 2019. Adaptive Designs for Clinical Trials of Drugs and Biologics Guidance for Industry. FDA Maryland.

Follmann, D., 1995. Multivariate tests for multiple endpoints in clinical trials. Statistics in medicine 14, 1163–1175.

Follmann, D., 1996. A simple multivariate test for one-sided alternatives. Journal of the American Statistical Association 91, 854–861.

Forsythe, A.B., 1987. Validity and power of tests when groups have been balanced for prognostic factors. Computational Statistics & Data Analysis 5, 193–200.

Friedman, L.M., Furberg, C., DeMets, D.L., 2010. Fundamentals of clinical trials. volume 4. Springer.

van Gijn, W., Marijnen, C.A., Nagtegaal, I.D., Kranenbarg, E.M.K., Putter, H., Wiggers, T., Rutten, H.J., Påhlman, L., Glimelius, B., van de Velde, C.J., et al., 2011. Preoperative radiotherapy combined with total mesorectal excision for resectable rectal cancer: 12-year follow-up of the multicentre, randomised controlled tme trial. The lancet oncology 12, 575–582.

Gwise, T., Hu, F., Hu, J., 2008. Optimal biased coins for two-arm clinical trials. Statistics and its Interface 1, 125–135.

Hagino, A., Hamada, C., Yoshimura, I., Ohashi, Y., Sakamoto, J., Nakazato, H., 2004.

Statistical comparison of random allocation methods in cancer clinical trials. Controlled clinical trials 25, 572–584.

Heritier, S., Gebski, V., Pillai, A., 2005. Dynamic balancing randomization in controlled clinical trials. Statistics in medicine 24, 3729–3741.

Hu, F., Hu, Y., Ma, W., Zhang, L., Zhu, H., 2015. Statistical inference of adaptive randomized clinical trials for personalized medicine. Clinical Investigation 5, 415–425.

Hu, F., Rosenberger, W.F., 2003. Optimality, variability, power: evaluating response-adaptive randomization procedures for treatment comparisons. Journal of the American Statistical Association 98, 671–678.

Hu, F., Rosenberger, W.F., 2006. The theory of response-adaptive randomization in clinical trials. volume 525. John Wiley & Sons.

Hu, F., Rosenberger, W.F., Zhang, L.X., 2006. Asymptotically best response-adaptive randomization procedures. Journal of Statistical Planning and Inference 136, 1911–1922.

Hu, F., Zhang, L.X., 2004. Asymptotic properties of doubly adaptive biased coin designs for multitreatment clinical trials. Annals of Statistics 32, 268–301.

Hu, F., Zhang, L.X., He, X., 2009. Efficient randomized-adaptive designs. The Annals of Statistics , 2543–2560.

Hu, Y., Hu, F., et al., 2012. Asymptotic properties of covariate-adaptive randomization. The Annals of Statistics 40, 1794–1815.

Ivanova, A., 2003. A play-the-winner-type urn design with reduced variability. Metrika 58, 1–13.

Ivanova, A., Rosenberger, W.F., 2000. A comparison of urn designs for randomized clinical trials of k¿ 2 treatments. Journal of biopharmaceutical statistics 10, 93–107.

Ivanova, A., Rosenberger, W.F., 2001. Adaptive designs for clinical trials with highly successful treatments. Drug information journal 35, 1087–1093.

Jennison, C., Turnbull, B.W., 1993. Group sequential tests for bivariate response: interim analyses of clinical trials with both efficacy and safety endpoints. Biometrics , 741–752.

Jennison, C., Turnbull, B.W., 1999. Group sequential methods with applications to clinical trials. CRC Press.

Jennison, C., Turnbull, B.W., 2000. Group Sequential Methods with Applications to Clinical Trials. Chapman and Hall/CRC, Boca Raton, FL.

Jeon, Y., Hu, F., 2010. Optimal adaptive designs for binary response trials with three treatments. Statistics in Biopharmaceutical Research 2, 310–318.

Jin, Z., Lin, D., Wei, L., Ying, Z., 2003. Rank-based inference for the accelerated failure time model. Biometrika 90, 341–353.

Kalbfleisch, J.D., Prentice, R.L., 2011. The statistical analysis of failure time data. volume 360. John Wiley & Sons.

Karush, W., 1939. Minima of functions of several variables with inequalities as side constraints. Master's thesis. Dept. of Mathematics, Univ. of Chicago.

Koul, H., Susarla, V.v., Van Ryzin, J., et al., 1981. Regression analysis with randomly right-censored data. The Annals of statistics 9, 1276–1288.

Kuhn, H.W., Tucker, A.W., 1951. Nonlinear programming, in: Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, Berkeley, Calif.. pp. 481–492.

Lachin, J.M., 1988. Properties of simple randomization in clinical trials. Controlled clinical trials 9, 312–326.

Lai, T.L., Ying, Z., 1991. Large sample theory of a modified buckley-james estimator for regression analysis with censored data. The Annals of Statistics , 1370–1402.

Lal, B.K., Beach, K.W., Roubin, G.S., Lutsep, H.L., Moore, W.S., Malas, M.B., Chiu, D., Gonzales, N.R., Burke, J.L., Rinaldi, M., et al., 2012. Restenosis after carotid artery stenting and endarterectomy: a secondary analysis of crest, a randomised controlled trial. The Lancet Neurology 11, 755–763.

Lauter, J., 1996. Exact t and f tests for analyzing studies with multiple endpoints. Biometrics , 964–970.

Läuter, J., Glimm, E., Kropf, S., et al., 1998. Multivariate tests based on left-spherically distributed linear scores. The Annals of Statistics 26, 1972–1988.

Lebowitsch, J., Ge, Y., Young, B., Hu, F., 2012. Generalized multidimensional dynamic allocation method. Statistics in medicine 31, 3537–3544.

Lefkopoulou, M., Ryan, L., 1993. Global tests for multiple binary outcomes. Biometrics , 975–988.

Leon, L.F., Cai, T., Wei, L., 2009. Robust inferences for covariate effects on survival time with censored linear regression models. Statistics in Biosciences 1, 50–64.

Lin, D., Ying, Z., 1995. Semiparametric inference for the accelerated life model with time-dependent covariates. Journal of statistical planning and inference 44, 47–63.

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

Matts, J.P., Lachin, J.M., 1988. Properties of permuted-block randomization in clinical trials. Controlled clinical trials 9, 327–344.

Melfi, V., Page, C., 1998. Variablility in adaptive designs for estimation of success probabilities. Lecture Notes-Monograph Series , 106–114.

Melfi, V.F., Page, C., Geraldes, M., 2001. An adaptive randomized design with application to estimation. Canadian Journal of Statistics 29, 107–116.

Miller, R., Halpern, J., 1982. Regression with censored data. Biometrika 69, 521–531.

Moyé, L.A., 2003. Multiple analyses in clinical trials: fundamentals for investigators. Springer Science & Business Media.

Nocturnal Oxygen Therapy Trial Group, 1980. Continuous or nocturnal oxygen therapy in hypoxemic chronic obstructive lung disease: A clinical trial. Ann Intern Med 93, 391–398.

Nordle, O., Brantmark, B., 1977. A self-adjusting randomization plan for allocation of patients into two treatment groups. Clinical pharmacology and therapeutics 22, 825–830.

O'Brien, P.C., 1984. Procedures for comparing samples with multiple endpoints. Biometrics , 1079–1087.

Orbe, J., Ferreira, E., Núñez-Antón, V., 2002. Comparing proportional hazards and accelerated failure time models for survival analysis. Statistics in medicine 21, 3493–3510.

Palmer, C.R., 1991. A comparative phase ii clinical trials procedure for choosing the best of three treatments. Statistics in Medicine 10, 1327–1340.

Patnaik, P., 1949. The non-central χ 2-and f-distribution and their applications. Biometrika 36, 202–232.

Pocock, S.J., Geller, N.L., Tsiatis, A.A., 1987. The analysis of multiple endpoints in clinical trials. Biometrics , 487–498.

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

Ponnuraja, C., Venkatesan, P., 2010. Survival models for exploring tuberculosis clinical trial data-an empirical comparison. Indian Journal of Science and Technology 3, 755–758.

Quintaliani, G., Reboldi, G., Di Napoli, A., Nordio, M., Limido, A., Aucella, F., Messa, P., Brunori, G., 2020. Exposure to novel coronavirus in patients on renal replacement therapy during the exponential phase of covid-19 pandemic: survey of the italian society of nephrology. Journal of Nephrology , 1–12.

Ritov, Y., 1990. Estimation in a linear regression model with censored data. The Annals of Statistics , 303–328.

Rittmeyer, A., Barlesi, F., Waterkamp, D., Park, K., Ciardiello, F., von Pawel, J., Gadgeel, S.M., Hida, T., Kowalski, D.M., Dols, M.C., et al., 2017. Atezolizumab versus docetaxel in patients with previously treated non-small-cell lung cancer (oak): a phase 3, open-label, multicentre randomised controlled trial. The Lancet 389, 255–265.

Robbins, H., 1952. Some aspects of the sequential design of experiments. Bull. Amer. Math. Soc. 58, 527–535.

Roberts, R., Croft, C., Gold, H.K., Hartwell, T.D., Jaffe, A.S., Muller, J.E., Mullin, S.M., Parker, C., Passamani, E.R., Poole, W.K., et al., 1984. Effect of propranolol

on myocardial-infarct size in a randomized blinded multicenter trial. New England Journal of Medicine 311, 218–225.

Rosenberger, W.F., 2002. Randomized urn models and sequential design .

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

Rosenberger, W.F., Lachin, J.M., 2002. Randomization and the clinical trial. Randomization in Clinical Trials: Theory and Practice , 1–14.

Rosenberger, W.F., Lachin, J.M., 2015. Randomization in clinical trials: theory and practice. John Wiley & Sons.

Rosenberger, W.F., Stallard, N., Ivanova, A., Harper, C.N., Ricks, M.L., 2001. Optimal adaptive designs for binary response trials. Biometrics 57, 909–913.

Rosenberger, W.F., Sverdlov, O., 2008. Handling covariates in the design of clinical trials. Statistical Science , 404–419.

Rosenberger, W.F., Sverdlov, O., et al., 2008. Handling covariates in the design of clinical trials. Statistical Science 23, 404–419.

Rout, C., Rocke, D., Levin, J., Gouws, E., Reddy, D., 1993. A reevaluation of the role of crystalloid preload in the prevention of hypotension associated with spinal anesthesia for elective cesarean section. Anesthesiology: The Journal of the American Society of Anesthesiologists 79, 262–269.

Russell, D., Hoare, Z., Whitaker, R., Whitaker, C., Russell, I., 2011. Generalized method for adaptive randomization in clinical trials. Statistics in medicine 30, 922–934.

Senn, S., Bretz, F., 2007. Power and sample size when multiple endpoints are considered. Pharmaceutical Statistics 6, 161–170.

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

Signorini, D., Leung, O., Simes, R., Beller, E., Gebski, V., Callaghan, T., 1993. Dynamic balanced randomization for clinical trials. Statistics in medicine 12, 2343–2350.

Smith, R.L., 1984. Properties of biased coin designs in sequential clinical trials. The Annals of Statistics , 1018–1034.

Stute, W., 1993. Consistent estimation under random censorship when covariables are present. Journal of Multivariate Analysis 45, 89–103.

Stute, W., 1996. Distributional convergence under random censorship when covariables are present. Scandinavian journal of statistics , 461–471.

Tang, D.I., Geller, N.L., Pocock, S.J., 1993. On the design and analysis of randomized clinical trials with multiple endpoints. Biometrics , 23–30.

Tang, D.I., Gnecco, C., Geller, N.L., 1989a. An approximate likelihood ratio test for a normal mean vector with nonnegative components with application to clinical trials. Biometrika 76, 577–583.

Tang, D.I., Gnecco, C., Geller, N.L., 1989b. Design of group sequential clinical trials with multiple endpoints. Journal of the American Statistical Association 84, 775–779.

Taves, D.R., 1974. Minimization: a new method of assigning patients to treatment and control groups. Clinical Pharmacology & Therapeutics 15, 443–453.

Taves, D.R., 2010. The use of minimization in clinical trials. Contemporary clinical trials 31, 180–184.

Thompson, W.R., 1933. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. Biometrika 25, 285–294.

Toorawa, R., Adena, M., Donovan, M., Jones, S., Conlon, J., 2009. Use of simulation to compare the performance of minimization with stratified blocked randomization. Pharmaceutical Statistics: The Journal of Applied Statistics in the Pharmaceutical Industry 8, 264–278.

Tsai, K.T., Kozial, J.A., 1994. Assessing multiple endpoints with ordered alternatives. Communications in Statistics-Theory and Methods 23, 533–544.

Tsiatis, A.A., 1990. Estimating regression parameters using linear rank tests for censored data. The Annals of Statistics , 354–372.

Tymofyeyev, Y., Rosenberger, W.F., Hu, F., 2007. Implementing optimal allocation in sequential binary response experiments. Journal of the American Statistical Association 102, 224–234.

Urokinase Pulmonary Embolism Trial Study Group, 1974. Urokinase-streptokinase embolism trial: Phase 2 results. Journal of the American Medical Association 229, 1606–1613.

Van Der Graaf, W.T., Blay, J.Y., Chawla, S.P., Kim, D.W., Bui-Nguyen, B., Casali, P.G., Schöffski, P., Aglietta, M., Staddon, A.P., Beppu, Y., et al., 2012. Pazopanib for metastatic soft-tissue sarcoma (palette): a randomised, double-blind, placebo-controlled phase 3 trial. The Lancet 379, 1879–1886.

Vestbo, J., Anderson, J.A., Brook, R.D., Calverley, P.M., Celli, B.R., Crim, C., Martinez, F., Yates, J., Newby, D.E., Investigators, S., et al., 2016. Fluticasone furoate and vilanterol and survival in chronic obstructive pulmonary disease with heightened cardiovascular risk (summit): a double-blind randomised controlled trial. The Lancet 387, 1817–1826.

Wei, L., 1978. An application of an urn model to the design of sequential controlled clinical trials. Journal of the American Statistical Association 73, 559–563.

Wei, L., Durham, S., 1978. The randomized play-the-winner rule in medical trials. Journal of the American Statistical Association 73, 840–843.

Wei, L., Lachin, J.M., 1988. Properties of the urn randomization in clinical trials. Controlled clinical trials 9, 345–364.

Wei, L., et al., 1979. The generalized polya's urn design for sequential medical trials. The Annals of Statistics 7, 291–296.

Weir, C.J., Lees, K.R., 2003. Comparison of stratification and adaptive methods for treatment allocation in an acute stroke clinical trial. Statistics in Medicine 22, 705–726.

Yin, G., 2012. Clinical trial design: Bayesian and frequentist adaptive methods. volume 876. John Wiley & Sons.

Ying, Z., 1993. A large sample study of rank estimation for censored regression data. The Annals of Statistics , 76–99.

Zeiher, B.G., Artigas, A., Vincent, J.L., Dmitrienko, A., Jackson, K., Thompson, B.T., Bernard, G., Group, S.S., et al., 2004. Neutrophil elastase inhibition in acute lung injury: results of the strive study. Critical care medicine 32, 1695–1702.

Zelen, M., 1969. Play the winner rule and the controlled clinical trial. Journal of the American Statistical Association 64, 131–146.

Zhang, L., Rosenberger, W.F., 2006. Response-adaptive randomization for clinical trials with continuous outcomes. Biometrics 62, 562–569.

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

Zhao, W., 2014. A better alternative to stratified permuted block design for subject randomization in clinical trials. Statistics in medicine 33, 5239–5248.

Zhao, W., Weng, Y., 2011. Block urn design—a new randomization algorithm for sequential trials with two or more treatments and balanced or unbalanced allocation. Contemporary clinical trials 32, 953–961.

Zhu, H., Hu, F., 2009. Implementing optimal allocation for sequential continuous responses with multiple treatments. Journal of Statistical Planning and Inference 139, 2420–2430.

Zhu, H., Hu, F., 2012. Interim analysis of clinical trials based on urn models. Canadian Journal of Statistics 40, 550–568.

Zhu, H., Hu, F., et al., 2010. Sequential monitoring of response-adaptive randomized clinical trials. The Annals of Statistics 38, 2218–2241.