Response Adaptive Designs for Clinical Trials with Continuous Outcomes

by

Yu Zhang

A Thesis submitted to the Faculty of Graduate Studies of
The University of Manitoba
in partial fulfillment of the requirements of the degree of

Master of Science

Department of Statistics
University of Manitoba
Winnipeg, Manitoba

Abstract

Response adaptive designs are developed for ethical considerations, which sequentially modify the treatment allocations based on the accumulating information in the trial so more patients receive the potentially better treatment. Yi and Wang (2009) proposed a variance-penalized criterion to evaluate the performance of a response adaptive design based on the expected number of patients assigned to the better treatment and the power of statistical test. We use the variance-penalized criterion to examine different response adaptive randomization procedures for normally distributed responses. We propose a new target allocation proportion which increases the chance that more patients receive the better treatment. Simulation results indicate that our proposed design has the advantage of assigning more patients to the potentially better treatment with minimum loss in statistical power, and our design performs better than the designs in the literature based on the variance-penalized criterion.

Acknowledgments

I am very grateful to my supervisors Dr. Xikui Wang and Dr. Yanqing Yi whose encouragement, patience and guidance helped my research and writing of the thesis. I would like to thank committee members Dr. Saumen Mandal and Dr. Pappur Shivakumar for their helpful suggestions and kind support. I wish to thank the Faulty of Graduate Studies of the University of Manitoba for the Manitoba Graduate Scholarship.

I would like to give my special thanks to my family and my parents for their love, support and encouragement throughout my life. Lastly, I wish to express my gratitude to all of those who supported me in any respect during the completion of this thesis.

This thesis is dedicated to my parents
for their love, support
and encouragement.

Contents

Abstract i

Acknowledgments ii

Dedication iii

Contents iv

List of Tables vi

1 Introduction 1

1.1 Background 1

1.2 Response Adaptive Randomization 5

1.2.1 Ethical Issues 5

1.2.2 Types of Response Adaptive Designs 8

1.2.3 Statistical Challenges 12

1.3 About the Thesis 14

2 Response Adaptive Designs with Continuous Outcomes **18**

2.1 Response Adaptive Designs 19

2.1.1 The Doubly-Adaptive Biased Coin Design 19

2.1.2 Optimal Allocation Proportions 21

2.2 Relationship between Power and Variance 23

2.3 The Variance-Penalized Criterion 25

2.4 New Allocation Proportion 28

2.4.1 Formulation of the Problem 29

2.4.2 Comparison of Designs 31

3 Some Numerical Results **34**

3.1 Sample Size Determination 34

3.2 Simulations 37

3.3 Further Illustration 43

3.4 Summary 47

4 Conclusion **49**

Appendix **52**

Bibliography **69**

List of Tables

3.1 Simulated mean (standard deviation) of the allocation proportions to treatment A for normal responses in a trial with 10,000 replications. 39

3.2 Simulated mean of total responses for normal responses in a trial with 10,000 replications. 40

3.3 Simulated power for normal responses in a trial with 10,000 replications. 41

3.4 Simulated values of the variance-penalized mean for normal responses in a trial with 10,000 replications. 42

3.5 Simulated mean (standard deviation) of the allocation proportions to treatment A and power for normal responses with equal variances. 44

3.6 Simulated values of the expected total treatment successes (ETS) and variance-penalized mean (VPM) for normal responses with equal variances. 45

3.7 Simulated mean (standard deviation) of the allocation proportions to treatment A and power for normal responses. 46

3.8 Simulated values of the expected total treatment successes (ETS) and variance-penalized mean (VPM) for normal responses. 47

Chapter 1

Introduction

Response adaptive designs of clinical trials have recently attracted a lot of attention in the clinical and biostatistics literature. Developed for ethical considerations, response adaptive designs adapt the treatment allocation probabilities so that more patients in the trial will receive the potentially better treatment. However, the use of response adaptive designs in clinical trials induces dependency among the responses and makes current statistical procedures even more complicated, which brings great challenges to maintain the validity of the trials. This chapter introduces the background, the motivation as well as the main results of our research.

1.1 Background

Conducting a clinical trial is an important step in developing a new medical treatment for certain disease. The modern clinical trial can be traced back to the eighteenth century. In the past several decades, clinical trials have become the preferred method for the assessment of treatment efficacy. Advanced techniques of allocation

of treatments and implementation, and special methods of statistical analysis of clinical trials have been developed.

A clinical trial is basically an experiment designed to evaluate the efficacy of a new medical treatment or intervention in human beings. In most cases, a trial is referred to as a prospective comparison of two or more treatments, consisting of the new intervention(s), and one or more control treatment(s). The intervention techniques employed in a clinical trial may be a single or a combination of drugs, devices, or procedures. A new intervention is usually compared with the current standard treatment or a placebo control if the standard treatment is not available. The objective of any clinical trial is to provide a valid comparison of the different treatments. An important issue here is that if covariates which affect the outcome are not equally distributed between the treatment groups, then the comparison of the treatment effects may be biased. A key feature of the clinical trial is the mechanism of randomization, which provides the best method to achieve homogeneous treatment groups with respect to known or unknown covariates.

Clinical trials are conducted in a series of steps, which are called phases. From the development of a new drug or therapy to its approval, there is a phase I clinical trial, a small trial to examine drug tolerance with the objective of finding a suitable dose level which avoids unacceptable adverse side effects and a phase II clinical trial, a study to evaluate both efficacy and safety of the new drug or therapy. The usual phase III clinical trial is a therapeutic confirmatory study, in which treatment effectiveness is verified and the presence of long-term adverse side effects is monitored. A phase IV clinical trial is a therapeutic use study, which involves follow-up of patient status after the new drug or therapy has been approved for

general use. Although the clinical trials have been generally divided into four phases, sometimes a trial's purposes can overlap the definitions of the boundary, especially in phases II and III.

During the process of a clinical trial, phase III trial is the most rigorous and extensive part. In phase III studies, to compare treatment effects, patients are randomly assigned either to the new treatment or to the current standard treatment, or possibly placebo treatment. The method of randomly assigning patients to one of the study treatments, called randomization, is considered as the pivotal component of phase III clinical trials. Many phase III trials need a lot of patients and take a long time to give a statistically significant evaluation of the treatment efficacy.

Clinical trials are complicated, expensive, ethically challenging, and require great attention to planning and monitoring. A well conducted clinical trial provides the best way to evaluate the efficacy and safety of a new treatment. However, with the traditional clinical trials, their lack of flexibility and ethics has adversely affected clinical development, which leads to growing interest in adaptive designs. Due to significant efficiency, cost-savings and ethical advantages over traditional approaches, the use of adaptive design methods in clinical research and development has become very popular in recent years.

For clinical trials, adaptive designs would allow researchers to modify the trial at interim stages based on accrued data. The concept of adaptive design can be traced back to 1970s when the adaptive randomization for sequential clinical trials were introduced (Wei, 1978). In 2005, the Pharmaceutical Research and Manufacturers of America (PhRMA) established a working group to investigate strategies, methodologies and implementation of adaptive designs for regulatory consideration.

Given in an executive summary, the PhRMA working group referred to an adaptive design as "a clinical trial design that uses accumulating data to decide on how to modify aspects of the study as it continues, without undermining the validity and integrity of the trial" (Gallo et al.,2006). The purpose of the modifications is to make clinical trials more flexible, efficient and ethical.

In practice, adaptive designs for modifying on-going clinical trials, especially for early-phase studies and phase III trials, have been well developed and discussed. Based on adaptations applied, adaptive designs refer to as a general set of methods including sample size re-estimation, early stopping due to safety of efficacy, dropping inferior treatment groups, adaptive dose finding, adaptive randomization and so on. These methods not only identify clinical benefit of the study treatment in a more efficient way but also increase the probability of success in clinical trial. In addition, adaptive methods tend to minimize the enrollment of patients to unsafe or ineffective treatments, thus providing an ethical advantage for their use.

Although the efficiency and ethical advantages of adaptive designs seem to be very attractive, their validity and integrity are challenged. From statistical point of view, the modification of an on-going clinical trial based on accumulating data may introduce selection bias, which comes from the modification of treatment, allocation of treatment to patients, or early withdrawal of patients. In addition, the adaptation of a trial may lead to a totally different trial involving statistical tests which are inconsistent with the original hypotheses. These concerns will not only have an influence on accuracy of statistical inference for treatment effects, but also bring challenges for developing appropriate statistical methodology for adaptive clinical trials.

1.2 Response Adaptive Randomization

Randomization plays an important role in a clinical trial for the evaluation of a new treatment. However, traditional randomization with fixed allocation probabilities has raised great ethical controversies because patients in the trial will have a fixed chance of receiving the superior treatment. Developed alternatively for ethical considerations, response adaptive designs use the accumulating information from previous allocations and responses in the trial to sequentially modify the probability of treatment allocation, so that more patients will receive the potentially better treatment.

1.2.1 Ethical Issues

When designing a clinical trial, randomization has always been an essential issue. Traditional randomization is used to assign patients to different treatment groups in order to reduce the bias in statistical comparison of treatment effects. It provides a powerful method for comparing treatment effects and is employed in many fields of clinical research. Traditional balanced randomization has been regarded as the gold standard because it provides the best chance of balancing known or unknown covariates of the patients in different treatment groups. Furthermore, balanced randomization justifies the use of statistical methods for evaluating treatment effects and maximizes the statistical power of hypothesis testing. But balanced randomization also implies unethical allocation of treatments in the clinical study because half of the patients will eventually receive the inferior treatment in a trial with two treatments.

Clinical trials, as experiments on humans, involve complex ethical issues which are not encountered in many other scientific experiments. In a clinical study, new interventions, such as investigational drugs, are being tested for efficiency and safety. However, until a drug is proven to be effective and safe, or ineffective, even harmful, it is uncertain that which experimental treatment is better. This provides the ethical basis for including a patient in a clinical trial, which is known as clinical equipoise. Freedman, B. (1987) defined the standard of clinical equipoise as "there is genuine uncertainty within the expert medical community about the preferred treatment". When enrolled in a clinical trial, a patient is given a chance of being assigned to a potentially beneficial treatment, which may also be highly toxic. At the same time, the patient also has a chance to be assigned to a placebo, rather than being assigned to a treatment which may be proven to be very beneficial (or harmful) later. In this situation, how to inform patients who take part in a clinical trial highlights the main ethical problem in such a study, that is, the possible conflict between trying to ensure that each individual patient has the best chance of receiving the better treatment, and evaluating competing treatments as efficiently as possible so that future patients might benefit from the superior treatment identified from the study. These considerations present the delicate tension between individual ethics and collective ethics (Swartz and Lellouch, 1971) in clinical trials. Individual ethics consider what is best for the individual patients, while in collective ethics, the advancement of public health is paid more attention. Such ethical dilemma is controversial and it would be attractive to treat more patients by the potentially better treatment while maintaining the validity of the clinical research for evaluating treatment effects.

A good clinical trial should provide an appropriate balance between individual and collective ethics. However, equal randomization is used in most trials to balance the treatment assignments. The probability that a patient will receive the potentially better treatment in a trial of two treatments is only 1/2. Clearly from the ethical point of view, equal randomization is not justifiable in desperate medical situations. With traditional balanced randomization, the individual ethics is unjustifiably sacrificed for the benefit of collective ethics.

As an example, the ACTG 076 trial conducted by the AIDS clinical trial group (ACTG) is considered. The trial's purpose was to examine the effect of the drug anti-viral zidovudine (AZT) on reducing the risk of vertical HIV transmission from the infected mothers to their infants. Out of 476 pregnant women with HIV-infection enrolled in the trial, 238 of them were assigned to the AZT group by using permuted block randomization and the remaining 238 women were assigned to the placebo group. The data analysis reported in Connor et al. (1994) showed that the infection rate was 7.6% in the AZT group and 22.6% in the placebo group. The result indicated that the effect of AZT on reducing the HIV transmission is statistically significant, with 0.00006 being the P-value of testing the null hypothesis that there is no difference between the HIV-infection rates of the two treatment groups. The findings of the trial benefit the newborns of the HIV-infected women in the general population, but the randomization employed in the trial is controversial because a large number of infants in the placebo group were infected with HIV by this allocation rule. Yao and Wei (1996) claimed that a suitable adaptive design, such as randomized play-the-winner (RPW) design, could have been used to skew the allocation probability to favor the AZT group, resulting in fewer infected infants,

while simultaneously maintaining the efficiency of the test for the benefit of patients in the general population.

To alleviate the tension between the individual ethics and collective ethics, response adaptive designs are developed in the past several decades. The goal is to assign more patients in the trial to the potentially better treatment with minimal loss in the statistical power. Having clear advantage in ethics, response adaptive designs have become very popular and important in the clinical and biostatistics communities. With a response adaptive design, the treatment allocation probability is sequentially modified based on the accumulating information in the trial so that a fewer number of patients are assigned to the inferior treatment. The overall objective is to improve both the efficiency and ethics of the clinical trial without undermining the validity and integrity of the intended research. Therefore with such a design, the same conclusion as a trial with balanced randomization can be reached, and meanwhile a higher proportion of patients will receive the superior treatment. Based on medical ethics, Pullman and Wang (2001) argued that in desperate medical situations, the use of response adaptive design is not only ethically justifiable but morally required. This is largely because that in such situations, truly informed consents are not available. Consequently the use of the gold standard of balanced randomization becomes ethically infeasible.

1.2.2 Types of Response Adaptive Designs

The response adaptive design has been well studied in the literature, especially for binary responses. As the pioneering concept, Zelen (1969) proposed the play-the-winner (PW) rule, which achieved the ethical objective of skewing the allocation

probability in favor of the better treatment. Wei and Durham (1978) studied the randomized play-the-winner (RPW) procedure originated from the urn model in order to reduce selection bias. Eisele (1994) introduced the doubly-adaptive biased coin design (DBCD) which is a randomization procedure to target some specific allocation proportions. Melfi et al. (2001) considered the sequential maximum likelihood procedure, a special case of the doubly-adaptive biased coin design under the specific allocation function. Hu and Rosenberger (2003) provided a theoretical template for the comparison of response adaptive randomization procedures. Hu and Zhang (2004) improved Eisele's procedure and established the asymptotical normality of allocation proportions in the doubly-adaptive biased coin design.

There are two main classes of response adaptive designs proposed thus far. One is the design-driven response adaptive randomization, which is based on urn models. The procedure can be completely nonparametric, driven by intuition and the allocation is usually not optimal. The other one is the target-based response adaptive randomization which is called sequential estimation procedure. The design is based on a parametric model for the response variable and a target allocation which involves unknown parameters of that model. The target proportion is achieved by sequentially substituting updated estimates of those parameters.

A large class of response adaptive designs is based on urn models. Consider the situation where two independent treatments A and B with instantaneous binary outcomes (success/failure) are compared in a trial. Zelen (1969) introduced one of the first response adaptive designs, the play-the-winner procedure. For a sequence of entering patients the first one is treated by tossing a fair coin, i.e., with probability 1/2 to receive either treatment. Upon successful outcome, the subsequent patient is

assigned to the same treatment, while the patient is assigned to the other treatment following a failure. The play-the-winner rule takes into account all the previous patients' allocation and responses for the assignment of new patients to ensure that more of them will receive the potentially better treatment. However such a design introduces much selection bias and hence is not practical in clinical research.

As the randomized version of the play-the-winner rule, Wei and Durham (1978) introduced the randomized play-the-winner procedure, which can be implemented by adopting a suitable urn. Assume that an urn contains μ balls of type A and μ balls of type B representing treatments A and B respectively. When a patient arrives, a ball is drawn at random from the urn and then replaced. The corresponding treatment of this ball is applied to the patient. The structure of the urn is changed based on the response of a previous patient. If this response is a success, an additional α balls of the same type are put in the urn; otherwise, an additional β balls of the opposite type are put in the urn. In this way, the composition of the urn is changed to favor balls associated with the more successful treatment thus far. This process is denoted by RPW (μ, α, β) rule. It has been shown that on average the randomized play-the-winner procedure assigns more patients in the trial to the better treatment. Furthermore, this rule is less vulnerable to selection bias than the play-the-winner rule.

Ivanova (2003) proposed the drop-the-loser rule which is defined as follows. An urn initially contains three types of balls with at least one ball of each type. Balls of type A and type B represent treatments A and B respectively. Balls of type 0 are immigration balls. A ball is randomly drawn from the urn. If it is an immigration ball, no patient is treated, and the ball is returned to the urn together with one

type A ball and one type B ball. The procedure is repeated until a treatment ball is selected, then the corresponding treatment is assigned to the patient and the response is observed. If a failure occurs, the ball is not replaced. If a success occurs, the ball is replaced and the urn remains unchanged. Compared to many other response adaptive randomization procedures, the drop-the-loser rule has two basic advantages: it is fully randomized; it generates the minimum asymptotic variance of the allocation proportion in the class of designs which target the proportion $q_B/(q_B + q_A)$ (Hu, Rosenberger, and Zhang, 2006).

Other important families of urn models include the generalized Friedman's urn (also called generalized Polya's urn) (Wei, 1979, Rosenberger et al., 1997, Bai et al., 2002) and birth-and-death urn (Ivanova et al., 2000). The urn models and their properties are reviewed in Rosenberger (2002). In general, response adaptive designs based on urn models are easy to implement in clinical trials. They skew the allocation proportion in favor of the better treatment. However, they can only target a specific allocation proportion which is usually not optimal.

Another class of response adaptive designs is the sequential estimation procedures. Eisele (1994) and Eisele and Woodroofe (1995) introduced the doubly-adaptive biased coin design, a randomization procedure to target some allocation proportions. Unlike the designs based on urn models, the doubly-adaptive biased coin design is based on a parametric model for the response variable. The ideas of this design are inherited from those of Efron's biased coin design (Efron, 1971). However, the biased coin design depends only on the current proportion of patients assigned to each treatment, whereas the doubly-adaptive biased coin design also depends on the current estimate of the target allocation proportion. Melfi et al. (2001)

proposed a simpler procedure, the sequential maximum likelihood procedure, which is considered as a special case of the doubly-adaptive biased coin design.

With the modifications of Eisele's procedure, Hu and Zhang (2004) provided a general doubly-adaptive biased coin design for multitreatment allocation in a clinical trial. The desired allocation proportions are based on unknown parameters of the response model and are achieved by sequentially substituting updated estimates of those parameters. Studies by Hu and Zhang (2004) also showed an asymptotic normality of allocation proportions in the doubly-adaptive biased coin design. Rosenberger and Hu (2004) compared some response randomization procedures with binary outcomes and concluded that the doubly-adaptive biased coin design is the best asymptotically one to maintain power while targeting any specific allocation proportion.

1.2.3 Statistical Challenges

A response adaptive design is a data-dependent treatment allocation procedure which sequentially uses the accumulating information in the trial to modify the allocation probability for new patients. One statistical challenge is that response adaptive design generates dependent data which are difficult to analyze statistically. Another challenge is that response adaptive design creates high variability of treatment allocations which adversely affect the statistical power of hypothesis testing.

A real life example is the extracorporeal membrane oxygenation (ECMO) trial conducted by Bartlett and his colleagues (1985) at the University of Michigan. The randomized play-the-winner rule was used in this trial to compare treatment

effects of ECMO and conventional therapy. Based on historical studies, the ECMO technique appeared to have an probability of success of 0.75 for newborns with neonatal respiratory failure, while the probability of success is less than 0.2 with conventional therapy.

The randomized play-the-winner rule used in the Michigan ECMO trial can be described by an urn which contains one ball representing ECMO and one ball representing the conventional therapy. When a patient arrives, a ball is drawn at random and the corresponding treatment is applied to the patient. After the ball is replaced, the structure of the urn is changed based on the patient's response. If this response is a success, an additional ball of the same type is put in the urn; otherwise, an additional ball of the opposite type is put in the urn. In this way, the composition of the urn is changed to favor the balls associated with the more successful treatment. During the trial, the first infant was randomly assigned to ECMO and survived. The second infant received conventional treatment and died. Next eight infants were assigned to ECMO and all survived. Finally, two more infants were treated by ECMO and were successful. The trial was then terminated. In brief, a total of 12 infants were assigned, with only one infant assigned to conventional therapy and 11 assigned to ECMO. The only failure outcome was observed with conventional therapy. The data demonstrated a significant better performance of ECMO. However, the design and analysis of the Michigan ECMO trial are highly controversial. The randomized play-the-winner rule has a high variability of the allocation proportion especially for small sample in this trial. That is the reason why only one infant was allocated to the conventional therapy. In fact one observation was not sufficient to reflect the treatment effect. The goal of obtaining valid statistical inferences about

the effect of ECMO versus conventional therapy is not achieved by this design.

Developed alternatively for ethical concerns in clinical trials, response adaptive designs have the intended purpose of maximizing the number of patients receiving the currently better or more successful treatment. However, during the response adaptive randomization procedures, the adaptations of the treatment allocations create a dependency among the responses from the trial, which introduces extra variability when statistically testing the treatment effects. Hu and Rosenberger (2003) provided an explicit relationship between power and variability of a response adaptive design, which showed that additional variability might lead to significant losses of power. Therefore, when comparing different response adaptive designs, the variability of the procedure must be considered. A good response adaptive design should achieve the goal of assigning more patients to the better treatment with minimal loss in the power of the statistical test.

Another serious concern is that the response adaptive randomization is usually based on the assumption that the responses can be obtained immediately after treatment. However, it is common that the response is observed after a certain period of time but still during the recruitment period, thus the effect of delayed responses should be examined. Clearly with response adaptive methods, the data analysis will be more complex than that with traditional methods, which brings great statistical challenges to researchers.

1.3 About the Thesis

Response adaptive design has been developed to achieve a variety of clinical objectives and for different types of response variables. Statistical approaches and

methodologies also vary depending on study objectives and different natures of the responses. In the literature, various optimality criteria are proposed to achieve specific clinical objectives, such as minimizing the expected total number of treatment failures in the trial (when the response is either success or failure) or maximizing the expected total survival time (when the response is the patient's survival time after treatment). Types of responses investigated in the literature include binary responses (success or failure), continuous responses (such as normal) and survival times (subject to censoring when observed).

Specific statistical goals of response adaptive design include determining the optimal proportion of treatment allocations to best balance the individual ethics and collective ethics and developing a statistical rule of randomizing treatment allocations to specifically achieve the derived optimal proportion. Inevitably, the statistical rule of randomizing treatment allocations must adapt to past treatment allocations and patients' responses, hence response adaptive randomization. One popular rule to achieve a specific allocation proportion is the doubly-adaptive biased coin design developed by Hu and Zhang (2004). For determining optimal allocations, Jennison and Turnbull (2000) described a general procedure. Rosenberger et al. (2001) proposed an optimal allocation proportion for binary responses, which minimizes the expected total failures with a given power. Biswas and Mandal (2004) used the same approach to obtain a procedure which results in an optimal allocation for normal responses. The optimal allocation approach was also extended to survival trials by Zhang and Rosenberger (2007).

With the dual objectives of assigning more patients in the trial to the better treatment and minimizing the loss of power for statistical comparison, the tradi-

tional statistical approach is to maximize the expected total number of success subject to a constraint on the variability of the treatment allocation. Recently Yi and Wang (2009) proposed a variance-penalized criterion which combines both the expected total number of success and the variability of the treatment allocation into a single optimality objective. They demonstrated the superiority of this new criterion over other criteria in the literature. However, Yi and Wang (2009) considered only binary responses in their study.

The objective of our research is to extend the variance-penalized criterion to continuous outcomes. Specially, our research is focused on the normally distributed responses. We use the variance-penalized criterion to examine the optimality properties of different target allocation proportions proposed in the literature, including equal allocation, Neyman allocation and Biswas and Mandal (2004) allocation. We also proposed a new proportion of treatment allocation which increases the chance that more patients receive the better treatment. Simulation studies are carried out to compare our proposed design with the designs proposed in the literature.

Simulation results indicate that the variance-penalized criterion works well for evaluating the performance of a response adaptive design with continuous responses. It is also shown that our proposed design has the advantage of assigning more patients to the potentially better treatment with minimum loss in power of the statistical test. The performance of our design is better than the designs proposed in the literature under the variance-penalized criterion.

This thesis is organized as follows. Chapter 2 introduces the response adaptive randomization procedures as well as the variance-penalized criterion. A new treatment allocation proportion is proposed for comparing two treatments with normal

responses. In Chapter 3, simulations are carried out to evaluate the performances of our design and some other response adaptive designs proposed in the literature under the variance-penalized criterion. Conclusions and discussions are provided in Chapter 4.

Chapter 2

Response Adaptive Designs with Continuous Outcomes

Many clinical trials have continuous responses as the primary outcome. The doubly-adaptive biased coin design (Hu and Zhang, 2004) is a randomization procedure with the flexibility to target any desired allocation proportion for various types of responses. In most cases, it is desired to target an optimal allocation proportion which best balances the individual ethics and collective ethics.

Zhang and Rosenberger (2006) studied the response adaptive designs with normally distributed outcomes and evaluated several optimal allocation proportions by both theoretical and simulation results. Based on their results, Biswas and Mandal (2004) provided the best balance between ethical and power concern under their optimality criterion. In this chapter, we propose a new proportion of treatment allocation for normal responses and use the doubly-adaptive biased coin design to target the proportion. It is expected that our newly proposed design will further improve Biswas and Mandal's procedure. The variance-penalized criterion (Yi and Wang, 2009) was introduced to evaluate the overall performances of the new design

and other response adaptive designs proposed in the literature.

2.1 Response Adaptive Designs

Hu and Zhang (2004) proposed a general doubly-adaptive biased coin design which applies to the case where the desired allocation proportions are unknown, but estimated sequentially. This randomization procedure can target any specific allocation proportion, and includes sequential maximum likelihood procedure (Melfi et al., 2001) as a special case.

2.1.1 The Doubly-Adaptive Biased Coin Design

Response adaptive randomization procedures include two main families. One is based on urn models, such as the randomized play-the-winner procedure (Wei and Durham, 1978). The urn designs have the advantage of shifting the allocation probability to assign more patients to the better treatment. However, they only target a specific allocation proportion which is usually not optimal in a formal sense. In addition, they are not flexible to accommodate all types of responses. Only clinical trials with binary or multinomial outcomes can directly apply this kind of designs. Another class of response adaptive randomization procedures is based on sequential estimation, which has the flexibility to target any desired treatment allocation proportion in clinical trials with different types of responses. One popular rule of sequential estimation procedures is the doubly-adaptive biased coin design developed by Hu and Zhang (2004). The key component of the procedure is the

following allocation function $g(x, y)$ defined on $[0, 1] \times [0, 1]$ with a parameter $\gamma \geq 0$:

$$g(0, y) = 1,$$

$$g(1, y) = 0,$$

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

Unlike the urn designs, the doubly-adaptive biased coin design is based on a parametric model for the response variable, which involves an unknown parameter vector $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Let $\rho(\boldsymbol{\theta}) \in (0, 1)$ be a target allocation proportion. This proportion is sequentially estimated according to accumulative information in the trial. Consider the case of comparing two treatments A and B. To make the problem and the statistical analysis meaningful, assign the first patient to treatment A and the second one to treatment B. Suppose $j (j \geq 2)$ patients have been assigned to the treatments and their responses have been observed. The procedure will assign the $(j + 1)$th patient to treatment A with probability $g(j_A/j,\ \rho(\hat{\boldsymbol{\theta}}_j))$, where j_A/j is the current proportion of patients that have already been allocated to treatment A and $\hat{\boldsymbol{\theta}}_j$ is the maximum likelihood estimator of $\boldsymbol{\theta}$ based on data from the first j patients. Here we use the maximum likelihood estimator because as is shown in Yi and Wang (2008), the maximum likelihood estimator is asymptotically efficient in response adaptive designs. When $\gamma = 0$ and $\hat{\boldsymbol{\theta}}_j$ is the maximum likelihood estimator of $\boldsymbol{\theta}$, the allocation proportion for $(j + 1)$th patient will only depend on $\rho(\hat{\boldsymbol{\theta}}_j)$. This is a special case of the doubly-adaptive biased coin design, called the sequential maximum likelihood procedure (Melfi et al., 2001).

Note that the nonnegative parameter γ in the allocation function $g(x, y)$ affects

the variability of the randomization procedure. When $\gamma = 0$, the procedure has the highest variability. As γ goes to ∞, the procedure will achieve the lower bound of any randomization procedure targeting the same allocation proportion (Hu and Zhang, 2006). Therefore, γ should be carefully chosen to obtain a good tradeoff between the degree of randomization and the variability. It has been shown (Rosenberger and Hu, 2004) that the doubly-adaptive biased coin design with $\gamma = 2$ tends to have very good convergence for binary responses, and always has the same or slightly better power than equal randomization with fewer treatment failures.

2.1.2 Optimal Allocation Proportions

Response adaptive designs were introduced to achieve the ethical goal of assigning more patients in the clinical trial to the better treatment. Another important goal is to minimize the loss of power for treatment comparison. However, these two goals usually compete with each other and thus it is desirable to consider a treatment allocation proportion which is optimal in some sense. Most often optimal allocation proportions are determined on the basis of some specific optimality criteria. Jennison and Turnbull (2000) described a general procedure for determining optimal allocations.

In the case of binary outcomes, Rosenberger et al. (2001) proposed an optimal allocation proportion, which minimizes the expected total number of treatment failures under a variance constraint. Such optimality criterion generates an allocation proportion which is ethically favorable while preserving the power of statistical test.

Biswas and Mandal (2004) generalized the binary optimal allocation to continuous responses. Specially, they studied normally distributed responses and took into

account both the locations and variabilities of the responses distributions. Suppose that patients enter sequentially into a clinical trial with two treatments A and B under comparison. The corresponding responses with two treatments are normally distributed, say $X_A \sim N(\mu_A, \sigma_A^2)$ and $X_B \sim N(\mu_B, \sigma_B^2)$ respectively. Let n_A and n_B be the number of patients assigned to treatments A and B. The total sample size is $n = n_A + n_B$. Assume that a smaller response is more desirable to patients. Consequently a sufficiently large response indicates unfavorable response of the treatment, which is treated as a failure. Biswas and Mandal (2004) considered the following optimization problem:

$$\min \left\{ n_A \Phi \left(\frac{\mu_A - c}{\sigma_A} \right) + n_B \Phi \left(\frac{\mu_B - c}{\sigma_B} \right) \right\}, \tag{2.1}$$

subject to

$$\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B} = K, \tag{2.2}$$

where K is some constant and c is a threshold value for dichotomizing the responses. The restriction on $\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}$ will preserve a specified level of power for the test of treatment equivalence. In (2.1), Biswas and Mandal considered minimization of the total number of patients with response larger than a constant c. Since a smaller response is desirable, the response larger than c is treated as a failure. The minimization can be interpreted as minimization of the expected total failures. Solving this problem yields the corresponding allocation proportion

$$\rho = \frac{\sigma_A \sqrt{\Phi \left(\frac{\mu_B - c}{\sigma_B} \right)}}{\sigma_A \sqrt{\Phi \left(\frac{\mu_B - c}{\sigma_B} \right)} + \sigma_B \sqrt{\Phi \left(\frac{\mu_A - c}{\sigma_A} \right)}}. \tag{2.3}$$

In Biswas and Mandal's design, they used sequential maximum likelihood procedure to target this specific allocation. The design is implemented by sequentially substituting updated estimates of the parameters to find the allocation probability $\hat{\rho}$ to treatment A for the next patient.

2.2 Relationship between Power and Variance

Hu and Rosenberger (2003) provided a template to understand the explicit relationship between power and variability of a binary response adaptive design. Zhang and Rosenberger (2006) gave a solution to the problem with normal responses, which showed that additional variability might lead to loss of power.

Consider a fixed design with n_A and n_B independent normal responses. Further assume that the corresponding response $X_k \sim N(\mu_k, \sigma_k^2)$, $k = A, B$. I wish to test the hypothesis

$$H_0 : \mu_A - \mu_B = 0$$

versus the alternative hypothesis

$$H_a : \mu_A - \mu_B \neq 0.$$

The Wald test statistic is

$$Z = \frac{\hat{\mu}_A - \hat{\mu}_B}{\sqrt{\frac{\hat{\sigma}_A^2}{n_A} + \frac{\hat{\sigma}_B^2}{n_B}}}.$$

For $n_A \to \infty$ and $n_B \to \infty$, Z is asymptotically normally distributed $N(0, 1)$ and Z^2 is asymptotically chi-squared with one degree of freedom under H_0.

Under H_a, power can be expressed as an increasing function of the noncentrality parameter of the chi-squared distribution which is given by

$$\phi = \frac{(\mu_A - \mu_B)^2}{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}$$

$$= \frac{(\mu_A - \mu_B)^2}{\frac{\sigma_A^2}{n\rho + n(n_A/n - \rho)} + \frac{\sigma_B^2}{n(1-\rho) - n(n_A/n - \rho)}},$$

where $n = n_A + n_B$ and ρ is a fixed target allocation proportion.

Let $x = n_A/n - \rho$, define a function

$$f(x) = \frac{\phi(x)}{n} = \frac{(\mu_A - \mu_B)^2}{\frac{\sigma_A^2}{\rho + x} + \frac{\sigma_B^2}{1 - \rho - x}}.$$

Hu and Rosenberger (2003) introduced the Taylor expansion,

$$f(x) = f(0) + f'(0)x + f''(0)x^2/2 + o(x^2)$$

and obtained

$$\phi/n = \frac{(\mu_A - \mu_B)^2}{\frac{\sigma_A^2}{\rho} + \frac{\sigma_B^2}{1-\rho}} + (\mu_A - \mu_B)^2 \frac{\sigma_A^2(1-\rho)^2 - \sigma_B^2\rho^2}{(\sigma_A^2(1-\rho) + \sigma_B^2\rho)^2}(\frac{n_A}{n} - \rho)$$

$$- (\mu_A - \mu_B)^2 \frac{\sigma_A^2\sigma_B^2}{(\sigma_A^2(1-\rho) + \sigma_B^2\rho)^3}(\frac{n_A}{n} - \rho)^2$$

$$+ o((\frac{n_A}{n} - \rho)^2).$$

In response adaptive designs, let $N_k(n)$ be the number of patients allocated to treatment k after n patients have been treated in the clinical trial, $k = A, B$, then $N_A(n)$ and $N_B(n)$ are random. But the test statistic Z, with $N_A(n)$ and $N_B(n)$

substituted for n_A and n_B respectively, still has an asymptotic normal distribution. The noncentrality parameter becomes a random variable, thus the expectation of $\phi(N_A(n)/n)$ is considered. Since it has been shown that $N_A(n)/n$ is asymptotically unbiased for ρ (Hu and Zhang, 2004), we can assume that $E(N_A(n)/n - \rho) = 0$ for large n. Therefore, the average power of the randomization procedure for a large sample is a function of

$$-(\mu_A - \mu_B)^2 \frac{\sigma_A^2 \sigma_B^2}{(\sigma_A^2(1-\rho) + \sigma_B^2 \rho)^3} E(\frac{N_A(n)}{n} - \rho)^2, \tag{2.4}$$

which is a direct function of $Var(N_A(n)/n)$ in a design.

2.3 The Variance-Penalized Criterion

With the dual objectives of assigning more patients in the trial to the better treatment and minimizing the loss of power for statistical comparison, the traditional statistical approach is to maximize the expected total number of success subject to a constraint on the variability of the treatment allocation. Recently Yi and Wang (2009) proposed a variance-penalized criterion which combines both the expected total number of success and the variability of the treatment allocation into a single optimality objective.

Consider a clinical trial in which patients are sequentially allocated to one of the two competing treatments A and B. Assume that responses under treatment k, $k = A, B$, are independent and follow a distribution $f_k(x, \boldsymbol{\theta}_k)$, where the unknown parameter $\boldsymbol{\theta}_k$ may be labeled by the pair of the location parameter μ_k and the scale

parameter σ_k. Usually μ_k is the mean and σ_k^2 is the variance corresponding to the treatment k, if they exist.

Denote by X_{iA} or X_{iB} the response for the ith patient depending on the allocated treatment A or B. Let δ_i be the indicator of treatment allocation following the response adaptive randomization procedure such that $\delta_i = 1$ if the ith patient is allocated to treatment A, $\delta_i = 0$ if the ith patient is allocated to treatment B. Represent the entire sequence of observations by $\{Y_i\}$, then $Y_i = \delta_i X_{iA} + (1-\delta_i)X_{iB}$ for $i = 1, 2, \cdots, n, \cdots$.

In a response adaptive randomization procedure, the treatment allocation indicator δ_n for the nth patient depends on the accumulating information from past treatment allocations and previous responses $\{(\delta_1, y_1), \cdots, (\delta_{n-1}, y_{n-1})\}$, $n \geq 2$. The response adaptive design is specified by its allocation rule $\pi = \{\pi_n, n = 1, 2, \cdots\}$, which consists of a sequence of probabilities such that $\pi_n = P(\delta_n = 1 \mid (\delta_1, y_1), \cdots, (\delta_{n-1}, y_{n-1}))$, $n \geq 2$ and $\pi_1 = P(\delta_1 = 1)$ is a pre-fixed value.

As indicated earlier, a good response adaptive design is expected to assign as many patients as possible to the better treatment, with less variability. Based on the dual objectives, the variance-penalized criterion considers an objective function which combines both the expected total responses and a positive multiple of its variance. Assume that a larger response is desirable in a clinical trial, the objective is to maximize

$$\left\{ E\left(\sum_{i=1}^{n} Y_i\right) - \lambda Var\left(\sum_{i=1}^{n} Y_i\right) \right\}, \tag{2.5}$$

which is called the variance-penalized mean (Yi and Wang, 2009), where $\lambda > 0$ is the penalty parameter.

In the case that a smaller response is desirable, similar as in Yi and Wang (2009), we propose to minimize

$$\left\{ E\left(\sum_{i=1}^{n} Y_i\right) + \lambda Var\left(\sum_{i=1}^{n} Y_i\right) \right\}. \tag{2.6}$$

Let $N_k = N_k(n)$ be the number of patients allocated to treatment k after n patients have been treated in the clinical trial, $k = A, B$. Then $N_A = \sum_{i=1}^{n} \delta_i$, $N_B = \sum_{i=1}^{n}(1-\delta_i) = n - N_A$, and

$$\begin{aligned}\sum_{i=1}^{n} Y_i &= \sum_{i=1}^{n} \{\delta_i X_{iA} + (1-\delta_i) X_{iB}\} \\ &= \sum_{i=1}^{n} \delta_i X_{iA} + \sum_{i=1}^{n} (1-\delta_i) X_{iB}.\end{aligned}$$

Yi and Wang (2009) showed that,

$$E\left(\sum_{i=1}^{n} Y_i\right) = (\mu_A - \mu_B) E(N_A) + n\mu_B, \tag{2.7}$$

and

$$Var\left(\sum_{i=1}^{n} Y_i\right) = (\sigma_A^2 - \sigma_B^2) E(N_A) + n\sigma_B^2 + (\mu_A - \mu_B)^2 Var(N_A), \tag{2.8}$$

where μ_k is the mean and σ_k^2 is the variance of the response corresponding to the kth treatment, $k = A, B$.

Therefore, an optimal response adaptive design is equivalent to achieve

$$\max\left\{ [\mu_A - \mu_B - \lambda(\sigma_A^2 - \sigma_B^2)] E(N_A) - \lambda(\mu_A - \mu_B)^2 Var(N_A) \right\}. \tag{2.9}$$

Similar results with this modified variance-penalized criterion, an optimal response adaptive design is equivalent to achieve

$$\min \left\{ [\mu_A - \mu_B + \lambda(\sigma_A^2 - \sigma_B^2)] E(N_A) + \lambda(\mu_A - \mu_B)^2 Var(N_A) \right\}. \tag{2.10}$$

For a design with a larger value of the objective function, the first term of the criterion indicates that the design is expected to assign a higher proportion of patients to the better treatment. The second term indicates that the design has a smaller variance of the treatment allocation proportion for a fixed total number of patients. According to the results of Hu and Rosenberger (2003), reduced variability leads to improved power. Therefore, a good design under this criterion tends to allocate more patients to the better treatment and increase the power of the test.

In the case of $\sigma_A = \sigma_B$, more weight is put on the expected number of patients assigned to the better treatment for a small value of λ. For a large value of λ, the power of the test is emphasized. Thus the term related to the power in the equivalent objective function is considered as a penalty and the penalty parameter λ provides a balance between the number of patients assigned to the better treatment and the power of the test.

2.4 New Allocation Proportion

In this section, we propose a treatment allocation proportion for normally distributed responses which takes into account both the locations and variabilities of the responses distributions. The doubly-adaptive biased coin design is used to target the proportion.

2.4.1 Formulation of the Problem

Consider a response adaptive design with normally distributed outcomes. To be specific, suppose that two treatments A and B with normally distributed responses $X_A \sim N(\mu_A, \sigma_A^2)$ and $X_B \sim N(\mu_B, \sigma_B^2)$ respectively are compared in a clinical trial. A specified number n of patients arrive sequentially to be allocated to one of two treatments. Further assume that the patients' responses are known before the next patient is enrolled. Let n_A and n_B be the number of patients assigned to treatments A and B, then $n_A + n_B = n$. Initially, assume that n_A, n_B are fixed. Consider the null hypothesis

$$H_0 : \mu_A - \mu_B = 0$$

versus the alternative hypothesis

$$H_a : \mu_A - \mu_B \neq 0.$$

We use the Wald test statistic

$$Z = \frac{\hat{\mu}_A - \hat{\mu}_B}{\sqrt{\frac{\hat{\sigma}_A^2}{n_A} + \frac{\hat{\sigma}_B^2}{n_B}}} \tag{2.11}$$

for treatment comparison, which is asymptotically normally distributed $N(0, 1)$. The objective is to choose the treatment as effectively as possible. Furthermore, a design which assigns more patients to the better treatment is preferred due to ethical consideration.

For binary responses, Yi and Wang (2009) proposed a design achieving the goal of assigning more patients to the better treatment with less variability. Let p_A be the probability of a success on treatment A and p_B be the success probability of

treatment B, with $q_A = 1 - p_A$ and $q_B = 1 - p_B$. Their design was to target the treatment allocation proportion

$$\rho = \frac{q_B + \epsilon \min\{q_A, q_B\}\text{sign}(q_B - q_A)}{q_A + q_B}, \tag{2.12}$$

where $0 \leq \epsilon \leq 1$. The doubly-adaptive biased coin design was used to target the desired ρ. Compared with other existing designs, this design has a better performance according to the variance-penalized criterion.

To extend the idea to the case of normally distributed responses, dichotomizing the responses by setting an appropriate threshold is considered. Assume that a smaller response is more desirable to patients in the trial. Consequently a sufficiently large response may indicate adverse effect of the treatment and may be regarded as a failure. As in Biswas and Mandal's (2004) design, a threshold constant c is chosen to be the boundary between treatment effectiveness and treatment failure. That is, a response larger than a constant c is considered a failure. Our proposed design is to target the treatment allocation proportion

$$\rho = \frac{\Phi\left(\frac{\mu_B - c}{\sigma_B}\right) + \epsilon \min\left\{\Phi\left(\frac{\mu_A - c}{\sigma_A}\right), \Phi\left(\frac{\mu_B - c}{\sigma_B}\right)\right\}\text{sign}\left(\Phi\left(\frac{\mu_B - c}{\sigma_B}\right) - \Phi\left(\frac{\mu_A - c}{\sigma_A}\right)\right)}{\Phi\left(\frac{\mu_A - c}{\sigma_A}\right) + \Phi\left(\frac{\mu_B - c}{\sigma_B}\right)}, \tag{2.13}$$

where $0 \leq \epsilon \leq 1$.

Since $\Phi\left(\frac{\mu_k - c}{\sigma_k}\right) = P(X_k \geq c)$ can be considered as the failure rate of treatment k, $k = A, B$, $\Phi\left(\frac{\mu_B - c}{\sigma_B}\right) - \Phi\left(\frac{\mu_A - c}{\sigma_A}\right) > 0$ indicates that treatment A has a potentially better performance than treatment B does. If $\epsilon = 0$, the target proportion becomes

$$\rho = \frac{\Phi\left(\frac{\mu_B - c}{\sigma_B}\right)}{\Phi\left(\frac{\mu_A - c}{\sigma_A}\right) + \Phi\left(\frac{\mu_B - c}{\sigma_B}\right)}, \tag{2.14}$$

which is inversely proportional to the corresponding failure rate. If $\epsilon > 0$, the target proportion ρ increases to $\frac{\Phi\left(\frac{\mu_B - c}{\sigma_B}\right) + \epsilon\Phi\left(\frac{\mu_A - c}{\sigma_A}\right)}{\Phi\left(\frac{\mu_A - c}{\sigma_A}\right) + \Phi\left(\frac{\mu_B - c}{\sigma_B}\right)}$ for $\Phi\left(\frac{\mu_B - c}{\sigma_B}\right) > \Phi\left(\frac{\mu_A - c}{\sigma_A}\right)$ or reduces to $\frac{\Phi\left(\frac{\mu_B - c}{\sigma_B}\right) - \epsilon\Phi\left(\frac{\mu_B - c}{\sigma_B}\right)}{\Phi\left(\frac{\mu_A - c}{\sigma_A}\right) + \Phi\left(\frac{\mu_B - c}{\sigma_B}\right)}$ for $\Phi\left(\frac{\mu_B - c}{\sigma_B}\right) < \Phi\left(\frac{\mu_A - c}{\sigma_A}\right)$. This means that a higher proportion of patients is expected to be allocated to the potentially better treatment as ϵ increases. However, when ϵ approaches 1, targeting the proposed proportion ρ will assign only a few number of patients to the inferior treatment, which makes it inadequate to draw a valid statistical conclusion about the treatment effect. Therefore, ϵ should be properly chosen to provide an appropriate balance between the individual ethics and collective ethics.

2.4.2 Comparison of Designs

Using doubly-adaptive biased coin design to target our proposed proportion, this randomization procedure is expected to increase the chance that more patients receive the better treatment while maintaining the power of the statistical test. Zhang and Rosenberger (2006) have discussed some other response adaptive randomization procedures for normal responses. Based on their study, the allocation procedure of Biswas and Mandal (2004) (referred to as BM) provided the best balance between ethical and power concern. To examine the optimality properties of our proposed

design and BM's procedure, the variance-penalized criterion is considered for normally distributed responses.

Simulation studies will be carried out to compare different randomization procedures, including both equal allocation and the adaptive allocation scheme. As discussed in Zhang and Rosenberger (2006), the following allocation proportions are considered:

1. Equal allocation

   Under the completely randomized rule, any incoming patient is assigned to either treatment with probability 1/2.

2. Neyman allocation $\quad \rho = \dfrac{\sigma_A}{\sigma_A + \sigma_B}$

   This allocation proportion is obtained from the optimization problem

   $$\min \left\{ \frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B} \right\},$$

   subject to

   $$n_A + n_B = n,$$

   which means that the variance of the estimated treatment difference is minimized subject to a fixed trial size $n = n_A + n_B$. The doubly-adaptive biased coin design is used to target the Neyman allocation proportion.

3. BM allocation

   $$\rho = \frac{\sigma_A \sqrt{\Phi\left(\frac{\mu_B - c}{\sigma_B}\right)}}{\sigma_A \sqrt{\Phi\left(\frac{\mu_B - c}{\sigma_B}\right)} + \sigma_B \sqrt{\Phi\left(\frac{\mu_A - c}{\sigma_A}\right)}}$$

This allocation proportion is based on minimizing the expected total failures while keeping the variance at a fixed level. The sequential maximum likelihood procedure is used to target this specific allocation.

4. New allocation

$$\rho = \frac{\Phi\left(\frac{\mu_B - c}{\sigma_B}\right) + \epsilon \min\left\{\Phi\left(\frac{\mu_A - c}{\sigma_A}\right), \Phi\left(\frac{\mu_B - c}{\sigma_B}\right)\right\} \text{sign}\left(\Phi\left(\frac{\mu_B - c}{\sigma_B}\right) - \Phi\left(\frac{\mu_A - c}{\sigma_A}\right)\right)}{\Phi\left(\frac{\mu_A - c}{\sigma_A}\right) + \Phi\left(\frac{\mu_B - c}{\sigma_B}\right)},$$

where $0 \leq \epsilon \leq 1$.

All these designs can be evaluated under the variance-penalized criterion, no matter what the target proportion is. Since a smaller response is more desirable to patients, the objective is to maximize

$$\left\{ - E\left(\sum_{i=1}^{n} Y_i\right) - \lambda Var\left(\sum_{i=1}^{n} Y_i\right) \right\}, \tag{2.15}$$

where $\lambda > 0$ is the penalty parameter.

Our proposed design will be compared with the above existing designs by means of simulation. The doubly-adaptive biased coin design is used to target the proposed allocation proportion.

Chapter 3

Some Numerical Results

In this chapter a detailed simulation study is presented to investigate the performance of the proposed design with normal responses, in comparison with some existing designs for clinical trials. The numerical results indicate that our proposed design is better than the response adaptive designs proposed in the literature based on the variance-penalized criterion.

3.1 Sample Size Determination

When planning a clinical trial, it is necessary to determine the total number of patients to be randomized. In practice, the number of patients involved in a study usually depends on the expense of data collection, and the need to have sufficient statistical power. In the simulation study, it is also desirable to determine a target sample size based on some model. Typically, the sample size is computed to provide a fixed level of power under a specified alternative hypothesis.

Suppose that two treatments A and B with normally distributed responses $X_A \sim N(\mu_A, \sigma_A^2)$ and $X_B \sim N(\mu_B, \sigma_B^2)$ respectively are to be tested for a statistically

significant difference between the population means, using the two-sided z-test. Consider the null hypothesis

$$H_0 : \mu_A - \mu_B = 0$$

versus the alternative hypothesis

$$H_a : \mu_A - \mu_B \neq 0.$$

When conducting a statistical test, two types of error must be considered: Type I error and Type II error, with probabilities α and β respectively. Power $(1-\beta)$ is the probability that the null hypothesis will be rejected if the alternative is true, which is an important consideration for determining sample size. Standard regulatory criteria for clinical trials often lead to specifying α to be 0.05 and power to be at least 0.8.

Assume that the two independent groups are of the same size n and that the pooled standard deviation of the population is taken as $\sigma = \sqrt{\frac{\sigma_A^2+\sigma_B^2}{2}}$. Then under the alternative hypothesis, $\bar{X}_A - \bar{X}_B \sim N(\mu_A - \mu_B, \frac{2\sigma^2}{n})$. The objective is to determine the desired sample size n, which simultaneously satisfies the conditions

$$P_r(|Z| \geq Z_{1-\frac{\alpha}{2}}|H_0) = \alpha$$

and

$$P_r(|Z| \geq Z_{1-\frac{\alpha}{2}}|H_a) = 1 - \beta$$

where $Z_{1-\frac{\alpha}{2}}$ is the standard normal deviate at the significance level $(1 - \frac{\alpha}{2})$. The basic relationship used to determine n, based on values of α and β under a specified

alternative hypothesis can be derived as follows:

$$
\begin{aligned}
1-\beta &= P_r(|Z| \geq Z_{1-\frac{\alpha}{2}}|H_a) \\
&= P_r\left(\left|\frac{\bar{X}_A-\bar{X}_B}{\sigma\sqrt{\frac{2}{n}}}\right| \geq Z_{1-\frac{\alpha}{2}} \mid H_a\right) \\
&= P_r\left(\frac{\bar{X}_A-\bar{X}_B}{\sigma\sqrt{\frac{2}{n}}} \geq Z_{1-\frac{\alpha}{2}} \mid H_a\right) + P_r\left(\frac{\bar{X}_A-\bar{X}_B}{\sigma\sqrt{\frac{2}{n}}} \leq -Z_{1-\frac{\alpha}{2}} \mid H_a\right) \\
&= P_r\left(\frac{\bar{X}_A-\bar{X}_B-(\mu_A-\mu_B)}{\sigma\sqrt{\frac{2}{n}}} \geq Z_{1-\frac{\alpha}{2}} - \frac{\mu_A-\mu_B}{\sigma\sqrt{\frac{2}{n}}} \mid H_a\right) \\
&+ P_r\left(\frac{\bar{X}_A-\bar{X}_B-(\mu_A-\mu_B)}{\sigma\sqrt{\frac{2}{n}}} \leq -Z_{1-\frac{\alpha}{2}} - \frac{\mu_A-\mu_B}{\sigma\sqrt{\frac{2}{n}}} \mid H_a\right) \\
&= P_r\left(Z \geq Z_{1-\frac{\alpha}{2}} - \frac{\mu_A-\mu_B}{\sigma\sqrt{\frac{2}{n}}}\right) + P_r\left(Z \leq -Z_{1-\frac{\alpha}{2}} - \frac{\mu_A-\mu_B}{\sigma\sqrt{\frac{2}{n}}}\right),
\end{aligned}
$$

where Z is a standard normal variate. This implies that

$$
Z_\beta \cong Z_{1-\frac{\alpha}{2}} - \frac{\mu_A-\mu_B}{\sigma\sqrt{\frac{2}{n}}}
$$

or

$$
Z_{1-\beta} \cong -Z_{1-\frac{\alpha}{2}} - \frac{\mu_A-\mu_B}{\sigma\sqrt{\frac{2}{n}}}.
$$

Simple algebra then leads to the equation

$$
n \cong \frac{2\sigma^2(Z_{1-\frac{\alpha}{2}}+Z_{1-\beta})^2}{(\mu_A-\mu_B)^2}. \tag{3.1}
$$

3.2 Simulations

All simulations are carried out in MATLAB. Four different target allocations, described in last section, are included in this study: equal allocation, Neyman allocation, BM allocation and the new allocation. The doubly-adaptive biased coin design is used to target the Neyman allocation and the new allocation proportion. The randomization procedure takes into account both of the current proportion of patients assigned to each treatment and the current estimate of the desired allocation proportion using the continuous observations.

Complete randomization which is not response adaptive is included in the study for the sake of comparison. In the simulations for all the adaptive designs, the allocation scheme assigns the first patient to treatment A and the second patient to treatment B. A random number from $N(\mu_A, \sigma_A^2)$ and $N(\mu_B, \sigma_B^2)$ each is drawn to denote the responses. Repeat until some initial estimates of (μ_A, σ_A) and (μ_B, σ_B) are obtained. The remaining patients are allocated by the designs. For each of the successive patients, the current maximum likelihood estimates $(\hat{\mu}_A, \hat{\sigma}_A)$ and $(\hat{\mu}_B, \hat{\sigma}_B)$ are used to determine the allocation.

Suppose $j(j \geq 2)$ patients have been assigned to treatments. When applying the doubly-adaptive biased coin design to target the allocation proportions, the probability of assigning $(j+1)$th patient to treatment A is given by $g(j_A/j,\ \hat{\rho})$, where g is an allocation function (Hu and Zhang, 2004):

$$g(x, y) = \frac{y(y/x)^\gamma}{y(y/x)^\gamma + (1-y)((1-y)/(1-x))^\gamma}. \tag{3.2}$$

Here j_A/j is the proportion of patients allocated to treatment A so far and $\hat{\rho}$, which

can be replaced by $(\hat{\mu}, \hat{\sigma})$, is the estimate of desired allocation proportion based on the first j patients. When $\gamma = 0$, $g(x, y) = y$, and this leads to the sequential maximum likelihood procedure, which is used to target BM allocation. After a fixed number n of patients have been thus allocated, the final allocation proportion is obtained. The total responses for all patients are also calculated. In addition, a test of the null hypothesis $H_0 : \mu_A = \mu_B$, the equality of treatment effects, against $H_a : \mu_A \neq \mu_B$ is performed.

After several trials and errors, $\epsilon = 0.3$ and $\epsilon = 0.5$ are considered for our proposed allocation. Following the original paper (Biswas and Mandal, 2004), parameter $c = 0$ is taken in both the BM design and our proposed design. As recommended in Rosenberger and Hu (2004), $\gamma = 2$ is used in the allocation function (3.2).

Each simulation is based on 10,000 replications. The combinations of parameters $(\mu_A, \sigma_A, \mu_B, \sigma_B)$ from Zhang and Rosenberger (2006) are used in the simulation. The sample size is chosen to yield simulated power of approximately 80 percent under balanced randomization, which is given by

$$\frac{4\sigma^2(Z_{1-\frac{\alpha}{2}} + Z_{1-\beta})^2}{(\mu_A - \mu_B)^2}, \text{ where } \sigma = \sqrt{\frac{\sigma_A^2 + \sigma_B^2}{2}}.$$

$\alpha = 0.05$ is taken in the two-sided test. The detailed simulation results are provided in Tables 3.1, 3.2, 3.3 and 3.4.

Table 3.1: Simulated mean (standard deviation) of the allocation proportions to treatment A for normal responses in a trial with 10,000 replications.

$(n,\ \mu_A,\ \mu_B,\ \sigma_A,\ \sigma_B)$	*Complete*	*Neyman*	*BM*	*New* ($\epsilon = 0.3$)	*New* ($\epsilon = 0.5$)
(88, 13, 15, 4.0, 2.5)	0.50	0.62	0.62	0.65	0.74
	(0.05)	(0.13)	(0.13)	(0.03)	(0.04)
(88, 13, 15, 2.5, 4.0)	0.50	0.37	0.38	0.37	0.30
	(0.05)	(0.13)	(0.13)	(0.06)	(0.12)
(88, 17, 15, 4.0, 2.5)	0.50	0.63	0.62	0.63	0.70
	(0.05)	(0.13)	(0.13)	(0.06)	(0.11)
(88, 17, 15, 2.5, 4.0)	0.50	0.37	0.38	0.35	0.26
	(0.05)	(0.13)	(0.13)	(0.03)	(0.05)
(350, 14, 15, 4.0, 2.5)	0.50	0.62	0.62	0.65	0.75
	(0.03)	(0.07)	(0.07)	(0.01)	(0.01)
(350, 14, 15, 2.5, 4.0)	0.50	0.38	0.38	0.35	0.25
	(0.03)	(0.07)	(0.07)	(0.01)	(0.01)
(350, 16, 15, 4.0, 2.5)	0.50	0.62	0.62	0.65	0.75
	(0.03)	(0.07)	(0.07)	(0.01)	(0.01)
(350, 16, 15, 2.5, 4.0)	0.50	0.38	0.38	0.35	0.25
	(0.03)	(0.07)	(0.07)	(0.01)	(0.01)

Table 3.1 gives the simulated values of the expected proportion of patients receiving treatment A along with the standard deviation. Compared with complete randomization, the result shows that response adaptive randomization skews the allocation probabilities so that more patients in the trial are assigned to the potentially better treatment. For normal responses, the determination of the better treatment depends on both parameters μ and σ. Our proposed design with $\epsilon = 0.5$ assigns higher expected proportions of patients to the potentially better treatment than all other designs. The proposed design with $\epsilon = 0.3$ has a better performance

in allocation proportion, with much smaller standard deviations than both the Neyman and BM designs. The simulated values of the expected total responses are also computed, which are given in Table 3.2.

Table 3.2: Simulated mean of total responses for normal responses in a trial with 10,000 replications.

$(n,\ \mu_A,\ \mu_B,\ \sigma_A,\ \sigma_B)$	*Complete*	*Neyman*	*BM*	*New* ($\epsilon = 0.3$)	*New* ($\epsilon = 0.5$)
(88, 13, 15, 4.0, 2.5)	1233	1210	1211	1206	1191
(88, 13, 15, 2.5, 4.0)	1232	1254	1253	1255	1267
(88, 17, 15, 4.0, 2.5)	1408	1430	1429	1431	1444
(88, 17, 15, 2.5, 4.0)	1407	1386	1387	1382	1366
(350, 14, 15, 4.0, 2.5)	5075	5034	5035	5021	4988
(350, 14, 15, 2.5, 4.0)	5075	5116	5119	5127	5161
(350, 16, 15, 4.0, 2.5)	5426	5467	5467	5478	5512
(350, 16, 15, 2.5, 4.0)	5424	5383	5383	5373	5338

For ethical consideration, more individuals in the response adaptive clinical trial should benefit from the potentially better treatment. In our setting, the expected total responses is used to represent this ethical constraint. Since a smaller response is preferred, the procedure giving the smallest expected total responses is considered as the most ethical procedure. It is found that our proposed design with $\epsilon = 0.3$ or $\epsilon = 0.5$ usually has a smaller expected total responses than both Neyman and BM procedures in the case that the allocation probabilities are skewed to the treatment with a smaller mean and a lager variance. However, when the allocation probabilities are skewed to the treatment with a larger mean and a lager variance, the expected total responses from the proposed design with $\epsilon = 0.5$ is larger than the other

designs. In this case, the proposed design with $\epsilon = 0.3$ has a similar performance as Neyman and BM designs.

Statistical power is another concern in the response adaptive design. The simulated power is computed for two-sided test of the equality of treatment effects. Table 3.3 gives the simulation results of power.

Table 3.3: Simulated power for normal responses in a trial with 10,000 replications.

$(n,\ \mu_A,\ \mu_B,\ \sigma_A,\ \sigma_B)$	*Complete*	*Neyman*	*BM*	*New* ($\epsilon = 0.3$)	*New* ($\epsilon = 0.5$)
(88, 13, 15, 4.0, 2.5)	0.79	0.82	0.80	0.81	0.78
(88, 13, 15, 2.5, 4.0)	0.79	0.81	0.80	0.82	0.79
(88, 17, 15, 4.0, 2.5)	0.79	0.81	0.80	0.81	0.79
(88, 17, 15, 2.5, 4.0)	0.79	0.81	0.80	0.81	0.78
(350, 14, 15, 4.0, 2.5)	0.80	0.81	0.81	0.83	0.79
(350, 14, 15, 2.5, 4.0)	0.79	0.82	0.82	0.82	0.79
(350, 16, 15, 4.0, 2.5)	0.80	0.82	0.82	0.82	0.77
(350, 16, 15, 2.5, 4.0)	0.80	0.81	0.82	0.82	0.78

Neyman allocation usually has a good performance in statistical power, and the result under our proposed design with $\epsilon = 0.3$ is very close to or even better than that of the Neyman design. The BM design has a slightly smaller power than these two designs. The proposed design with $\epsilon = 0.5$ has more loss in statistical power because of the larger skewing to the better treatment.

From Tables 3.1, 3.2 and 3.3, it is clear that the main objectives of the response adaptive design are fulfilled. A lager number of patients have been treated by the potentially better treatment with less loss in power of the statistical test. Our

proposed design with $\epsilon = 0.3$ has a better performance than both Neyman and BM designs when considering both individual benefits and collective ethics. To examine this result, the variance-penalized criterion is used. Since a smaller expected total responses is more desirable in the trial, the objective is to maximize

$$\left\{ -E\left(\sum_{i=1}^{n} Y_i\right) - \lambda Var\left(\sum_{i=1}^{n} Y_i\right) \right\}, \tag{3.3}$$

where Y_i is the response for the ith patient. Penalty parameter $\lambda = 0.5$ ia taken in the simulation.

Table 3.4: Simulated values of the variance-penalized mean for normal responses in a trial with 10,000 replications.

$(n,\ \mu_A,\ \mu_B,\ \sigma_A,\ \sigma_B)$	*Neyman*	*BM*	*New* ($\epsilon = 0.3$)	*New* ($\epsilon = 0.5$)
(88, 13, 15, 4.0, 2.5)	-2026	-2038	-1758	-1836
(88, 13, 15, 2.5, 4.0)	-2066	-2054	-1849	-2040
(88, 17, 15, 4.0, 2.5)	-2238	-2226	-2018	-2182
(88, 17, 15, 2.5, 4.0)	-2204	-2191	-1963	-1991
(350, 14, 15, 4.0, 2.5)	-7476	-7457	-7223	-7385
(350, 14, 15, 2.5, 4.0)	-7580	-7469	-7336	-7514
(350, 16, 15, 4.0, 2.5)	-7892	-7883	-7688	-7892
(350, 16, 15, 2.5, 4.0)	-7836	-7754	-7578	-7714

Table 3.4 shows that the variance-penalized mean is the largest under the proposed design with $\epsilon = 0.3$. As indicated earlier, this is due to higher proportion of patients assigned to the better treatment and meanwhile the statistical power remains good. The proposed design with $\epsilon = 0.5$ also has a larger variance-penalized mean than Neyman and BM designs in the case where the allocation probabilities

are skewed to the treatment with a smaller mean and a lager variance.

It is clear that parameter ϵ in the proposed design measures a tradeoff between the individual and collective ethics. The design with a larger value of ϵ will result in loss of statistical power while higher proportion of patients is assigned to the better treatment.

Note that the value of the constant c taken in the BM design and our design can have significant effect on the performance of the procedures. For example, if the standard deviations of normal responses are small, then a change in the value of c will lead to a much larger change in the allocation probabilities. However, there is no clear guideline on how to choose c.

From the simulation results, our proposed design with $\epsilon = 0.3$ gives a better balance between ethical and power concern than the other designs, and is considered to be the preferred procedure under the variance-penalized criterion. To further illustrate the optimality property of the proposed design, we also run simulations for the case where equal variances are assumed for normal responses.

3.3 Further Illustration

The idea of dichotomizing normal responses by setting a threshold c is used in both BM design and our design. Suppose that a small response is desirable, then a response larger than a threshold c is considered as a failure. Represent the entire sequence of responses by $\{Z_i\}$ such that $Z_i = 1$ if $Y_i \leq c$ and $Z_i = 0$ if $Y_i > c$, where Y_i is the response of the ith patient, $i = 1, 2, \cdots, n$.

The BM allocation is obtained by minimizing the expected total number of

responses larger than c, which is equivalent to maximizing $E(\sum_{i=1}^{n} Z_i)$ given n_A with a fixed level of power for the test of treatment equivalence. For our proposed design, a lager number of successes on average is expected as well as the good performance in statistical power.

Represent the variance-penalized criterion as

$$\max \left\{ E\left(\sum_{i=1}^{n} Z_i\right) - \lambda Var\left(\sum_{i=1}^{n} Z_i\right) \right\}, \tag{3.4}$$

where $\lambda > 0$ is the penalty parameter.

A simulation study of $n = 80$ is carried out to compare the BM design and our proposed design with $\epsilon = 0.3$ under this variance-penalized criterion. We assume that $\mu_A \leq \mu_B$. For the purpose of comparison, fix μ_A at 1.0 and varied μ_B from 1.0 at a interval of 0.2. Also, $\sigma_A = \sigma_B = 1$ is taken in the simulation, and γ, c, λ and α have the same setting as before. The simulation results are provided in Tables 3.5 and 3.6.

Table 3.5: Simulated mean (standard deviation) of the allocation proportions to treatment A and power for normal responses with equal variances.

	BM		*New*	
$\mu_B - \mu_A$	$E(\rho)$	*Power*	$E(\rho)$	*Power*
0.0	0.50(0.15)	0.05	0.50(0.14)	0.05
0.2	0.51(0.15)	0.13	0.57(0.13)	0.14
0.4	0.51(0.15)	0.40	0.62(0.10)	0.39
0.6	0.52(0.14)	0.71	0.65(0.07)	0.71
0.8	0.52(0.14)	0.90	0.66(0.05)	0.92
1.0	0.52(0.14)	0.96	0.67(0.04)	0.99

Table 3.6: Simulated values of the expected total treatment successes (ETS) and variance-penalized mean (VPM) for normal responses with equal variances.

	BM		*New*	
$\mu_B - \mu_A$	*ETS*	*VPM*	*ETS*	*VPM*
0.0	13	7	13	7
0.2	11	6	11	6
0.4	10	5	10	5
0.6	9	4	10	5
0.8	8	3	9	5
1.0	7	3	9	5

The result shows that our proposed procedure works better than BM procedure in terms of allocation proportion, power and expected total treatment successes. When the treatment difference is small or moderate, BM procedure places roughly half of the patients on each treatment group. Our proposed procedure skews the allocation probabilities when the treatment difference is present so that a larger number of patients are assigned to the better treatment. As the treatment difference increases, the allocation gets more skewed to the better treatment with the decrease in the standard deviation, and in the meanwhile the power increases quickly.

For the purpose of comparison, we also provide the simulation results of the case $\sigma_A \neq \sigma_B$. Two pairs of values of (σ_A, σ_B) are taken in the simulation. The simulation results are provided in Tables 3.7 and 3.8.

Table 3.7: Simulated mean (standard deviation) of the allocation proportions to treatment A and power for normal responses.

			BM		*New*	
$\mu_B - \mu_A$,	σ_A,	σ_B	$E(\rho)$	*Power*	$E(\rho)$	*Power*
0.0	2.0	1.0	0.70(0.13)	0.07	0.66(0.08)	0.05
0.2	2.0	1.0	0.70(0.13)	0.12	0.68(0.05)	0.09
0.4	2.0	1.0	0.71(0.13)	0.23	0.69(0.04)	0.21
0.6	2.0	1.0	0.71(0.13)	0.42	0.70(0.03)	0.42
0.8	2.0	1.0	0.71(0.13)	0.63	0.70(0.03)	0.64
1.0	2.0	1.0	0.71(0.13)	0.81	0.71(0.03)	0.83
0.0	1.0	2.0	0.30(0.14)	0.08	0.34(0.07)	0.05
0.2	1.0	2.0	0.31(0.13)	0.11	0.36(0.09)	0.09
0.4	1.0	2.0	0.31(0.13)	0.23	0.39(0.11)	0.21
0.6	1.0	2.0	0.32(0.13)	0.42	0.42(0.13)	0.41
0.8	1.0	2.0	0.32(0.13)	0.64	0.46(0.14)	0.64
1.0	1.0	2.0	0.33(0.14)	0.82	0.50(0.14)	0.81

Table 3.8: Simulated values of the expected total treatment successes (ETS) and variance-penalized mean (VPM) for normal responses.

			BM		*New*	
$\mu_B - \mu_A$,	σ_A,	σ_B	*ETS*	*VPM*	*ETS*	*VPM*
0.0	2.0	1.0	21	12	21	12
0.2	2.0	1.0	20	11	20	12
0.4	2.0	1.0	19	9	19	11
0.6	2.0	1.0	19	8	19	11
0.8	2.0	1.0	18	8	18	11
1.0	2.0	1.0	18	7	18	11
0.0	1.0	2.0	21	12	21	12
0.2	1.0	2.0	19	11	19	11
0.4	1.0	2.0	17	10	17	10
0.6	1.0	2.0	16	9	15	9
0.8	1.0	2.0	14	8	14	8
1.0	1.0	2.0	13	7	13	7

The modified variance-penalized criterion works well in evaluating the performance of these response adaptive randomization procedures. It is showed that the proposed design is better than BM design based on the variance-penalized criterion.

3.4 Summary

This chapter discusses the simulation results of several randomization procedures with a variance-penalized criterion. Our proposed design is compared with other existing designs according to the criterion. The overall performance of the proposed

design is good. It is showed that the proposed design has the potential to assign more patients to the better treatment in the trial but not necessarily result in a loss of statistical power.

Zhang and Rosenberger (2006) have evaluated equal, Neyman, BM and some other allocation proportions for normal responses by simulation study. Based on their results, BM allocation provided a better balance between ethical and power concern than the other adaptive allocations. It is expected that our proposed proportion of treatment allocation will further improve BM allocation's performance. Our simulation study shows that the proposed allocation results in a higher proportion of success with less variability than other allocations. BM allocation also assigns more patients to the better treatment, however, it has a slightly larger variance.

The variance-penalized criterion (Yi and Wang, 2009) combines both the mean and the variability of the total responses into a single optimality objective. The target is the expected value of total responses minus a positive multiple of its variance, called variance-penalized mean. The response adaptive design which has a larger value of variance-penalized mean is considered as a better design. From the simulation results, the performance of our proposed design is better than both Neyman and BM designs under the variance-penalized criterion. The proposed randomization procedure assigns more patients to the better treatment while the statistical power remains at a high level.

Chapter 4

Conclusion

Clinical trials are commonly employed in medical research for the assessment of new medical treatments. As experiment on human subjects, a clinical trial is typically characterized by the delicate tension between individual ethics and collective ethics. On one hand, patients in the trial should be given the best chance of receiving the better treatment, and on the other hand patients after the trials should benefit from the clinical study and receive the better treatment identified from the trial.

When designing a clinical trial, randomization has always been an essential feature, and equal randomization has traditionally been regarded as a gold standard for comparing treatment effects. But equal randomization also implies unethical allocation of treatments to patients because half of the patients in the trial will eventually receive the inferior treatment. From the ethical point of view, with equal randomization, the individual ethics is unjustifiably sacrificed for the benefit of collective ethics.

Response adaptive design of clinical trials is developed specifically to alleviate the tension between the individual ethics and collective ethics. In a response

adaptive design, the treatment allocations are sequentially modified based on the accumulating information of the treatment effect with the goal of assigning more patients to the better treatment. However, the high variability of treatment allocations will adversely affect the power for statistical comparison. Therefore, another important goal of a response adaptive design is to minimize the loss of power for treatment comparison.

In our research, we proposed a new allocation proportion for normal responses and applied the doubly-adaptive biased coin design to target this proportion. We also presented a simulation study to investigate the performance of our proposed design and some existing response adaptive designs. The simulation results indicated that response adaptive design has the ethical advantage of skewing the allocation probabilities to favor the better treatment. It was also showed that the unbalanced allocation induced by the response adaptive design does not severely sacrifice the power of the statistical test for comparing different treatments.

The overall performance of our proposed design was compared with other randomization procedures under a variance-penalized criterion. A good response adaptive design under this criterion is expected to assign more patients to the better treatment while maintaining a high level of power. Simulation results showed that the variance-penalized criterion works well for normal responses. Based on this criterion, our proposed design is better than other response adaptive designs.

Note that an underlying assumption to our research in response adaptive design is that previous patient's responses are available during the recruitment period. In practice, there might be a fixed duration of follow-up, especially in clinical trials with survival response. An allocation procedure for survival trials by using a parametric

approach that involves a target optimal allocation and a randomization procedure with low variability is waiting for exploration. Adjust the variance-penalized criterion to incorporate survival data and examine the effect of right censoring on the asymptotic behavior of the randomization procedures are under our future study.

Appendix

Computer Code

```
#Complete randomization

function [ETR, VTR, ERHO, SDRHO, POWER, y]
=complete(N, muA, muB, sigmaA, sigmaB)
MRHO=[];
MTR=[];
H=0;
lamda=0.5;
for t=1:1:10000
    n=0;
    nA=0;
    nB=0;
    rA=0;
    rB=0;
    NRNA=[];
    NRNB=[];
    for i=1:1:N
```

```
        n=n+1;
        urn=unifrnd(0,1,1);
        if urn<=0.5
           nA=nA+1;
           nrnA=normrnd(muA, sigmaA);
           NRNA=[NRNA nrnA];
           rA=rA+nrnA;
        else
           nB=nB+1;
           nrnB=normrnd(muB, sigmaB);
           NRNB=[NRNB nrnB];
           rB=rB+nrnB;
        end
    end
    NA=nA;
    NB=nB;
    RHO=NA/N;
    MRHO=[MRHO RHO];
    RA=rA;
    RB=rB;
    TR=RA+RB;
    MTR=[MTR TR];
    h=ttest2(NRNA, NRNB, [], [], 'unequal');
    H=H+h;
end

ERHO=mean(MRHO);
```

```
SDRHO=std(MRHO);
ETR=mean(MTR);
VTR=var(MTR);
POWER=H/10000;
y=-ETR-lamda*VTR;

#The Neyman design

function [ETR, VTR, ERHO, SDRHO, POWER, y]
=neyman(N, muA, muB, sigmaA, sigmaB)
MRHO=[];
MTR=[];
H=0;
lamda=0.5;
for t=1:1:10000
    n=0;
    nA=0;
    nB=0;
    rA=0;
    rB=0;
    NRNA=[];
    NRNB=[];
    SIGN=1;

    while SIGN==1
          n=n+1;
          nA=nA+1;
```

```
      nrnA=normrnd(muA, sigmaA);
      NRNA=[NRNA nrnA];
      rA=rA+nrnA;
      n=n+1;
      nB=nB+1;
      nrnB=normrnd(muB, sigmaB);
      NRNB=[NRNB nrnB];
      rB=rB+nrnB;
      sigmaAhat=std(NRNA)*sqrt((nA-1)/nA);
      sigmaBhat=std(NRNB)*sqrt((nB-1)/nB);
      sigmaThat=sigmaAhat+sigmaBhat;
      if sigmaThat==0
         SIGN=1;
      else
          SIGN=0;
      end
      tn=n+1;
end

gamma=2;

for i=tn:1:N
    x=nA/n;
    sigmaAhat=std(NRNA)*sqrt((nA-1)/nA);
    sigmaBhat=std(NRNB)*sqrt((nB-1)/nB);
    rouhat=sigmaAhat/(sigmaAhat+sigmaBhat);
    y=rouhat;
```

```
    g=(y*(y/x)^gamma)/(y*(y/x)^gamma+(1-y)*((1-y)/(1-x))^gamma);
    n=n+1;
    urn=unifrnd(0,1,1);
    if urn<=g
       nA=nA+1;
       nrnA=normrnd(muA, sigmaA);
       NRNA=[NRNA nrnA];
       rA=rA+nrnA;
    else
         nB=nB+1;
         nrnB=normrnd(muB, sigmaB);
         NRNB=[NRNB nrnB];
         rB=rB+nrnB;
    end
end

NA=nA;
NB=nB;
RHO=NA/N;
MRHO=[MRHO RHO];
RA=rA;
RB=rB;
TR=RA+RB;
MTR=[MTR TR];
h=ttest2(NRNA, NRNB, [], [], 'unequal');
H=H+h;
```

```
end

ERHO=mean(MRHO);

SDRHO=std(MRHO);

ETR=mean(MTR);

VTR=var(MTR);

POWER=H/10000;

y=-ETR-lamda*VTR;


#The BM design

function [ETR, VTR, ERHO, SDRHO, POWER, y]

=BM(N, muA, muB, sigmaA, sigmaB)

MRHO=[];

MTR=[];

H=0;

lamda=0.5;

for t=1:1:10000

    n=0;

    nA=0;

    nB=0;

    rA=0;

    rB=0;

    NRNA=[];

    NRNB=[];

    SIGN=1;
```

```
while SIGN==1
      n=n+1;
      nA=nA+1;
      nrnA=normrnd(muA, sigmaA);
      NRNA=[NRNA nrnA];
      rA=rA+nrnA;
      n=n+1;
      nB=nB+1;
      nrnB=normrnd(muB, sigmaB);
      NRNB=[NRNB nrnB];
      rB=rB+nrnB;
      sigmaAhat=std(NRNA)*sqrt((nA-1)/nA);
      sigmaBhat=std(NRNB)*sqrt((nB-1)/nB);
      if sigmaAhat==0 || sigmaBhat==0
         SIGN=1;
      else
          SIGN=0;
      end
      tn=n+1;
end

c=0;

for i=tn:1:N
    muAhat=mean(NRNA);
    muBhat=mean(NRNB);
    sigmaAhat=std(NRNA)*sqrt((nA-1)/nA);
```

```
    sigmaBhat=std(NRNB)*sqrt((nB-1)/nB);
    rouhat=sqrt(1-normcdf(c, muBhat, sigmaBhat))*sigmaAhat
    /(sqrt(1-normcdf(c, muBhat, sigmaBhat))*sigmaAhat
    +sqrt(1-normcdf(c, muAhat, sigmaAhat))*sigmaBhat);
    n=n+1;
    urn=unifrnd(0,1,1);
    if urn<=rouhat
       nA=nA+1;
       nrnA=normrnd(muA, sigmaA);
       NRNA=[NRNA nrnA];
       rA=rA+nrnA;
    else
         nB=nB+1;
         nrnB=normrnd(muB, sigmaB);
         NRNB=[NRNB nrnB];
         rB=rB+nrnB;
    end
end

NA=nA;
NB=nB;
RHO=NA/N;
MRHO=[MRHO RHO];
RA=rA;
RB=rB;
TR=RA+RB;
```

```
    MTR=[MTR TR];
    h=ttest2(NRNA, NRNB, [], [], 'unequal');
    H=H+h;
end

ERHO=mean(MRHO);
SDRHO=std(MRHO);
ETR=mean(MTR);
VTR=var(MTR);
POWER=H/10000;
y=-ETR-lamda*VTR;


#New design (epsilon=0.3)

function [ETR, VTR, ERHO, SDRHO, POWER, y]
=newdesign(N, muA, muB, sigmaA, sigmaB)
MRHO=[];
MTR=[];
H=0;
lamda=0.5;
for t=1:1:10000
    n=0;
    nA=0;
    nB=0;
    rA=0;
    rB=0;
```

```
NRNA=[];
NRNB=[];
SIGN=1;

while SIGN==1
      n=n+1;
      nA=nA+1;
      nrnA=normrnd(muA, sigmaA);
      NRNA=[NRNA nrnA];
      rA=rA+nrnA;
      n=n+1;
      nB=nB+1;
      nrnB=normrnd(muB, sigmaB);
      NRNB=[NRNB nrnB];
      rB=rB+nrnB;
      sigmaAhat=std(NRNA)*sqrt((nA-1)/nA);
      sigmaBhat=std(NRNB)*sqrt((nB-1)/nB);
      if sigmaAhat==0 || sigmaBhat==0
         SIGN=1;
      else
          SIGN=0;
      end
      tn=n+1;
end

gamma=2;
e=0.3;
```

```
c=0;

for i=tn:1:N
    x=nA/n;
    muAhat=mean(NRNA);
    muBhat=mean(NRNB);
    sigmaAhat=std(NRNA)*sqrt((nA-1)/nA);
    sigmaBhat=std(NRNB)*sqrt((nB-1)/nB);
    rouhat=(1-normcdf(c, muBhat, sigmaBhat)+
    e*min((1-normcdf(c, muAhat, sigmaAhat)),
    (1-normcdf(c, muBhat, sigmaBhat)))
    *sign((1-normcdf(c, muBhat, sigmaBhat))
    -(1-normcdf(c, muAhat, sigmaAhat))))
    /((1-normcdf(c, muAhat, sigmaAhat))
    +(1-normcdf(c, muBhat, sigmaBhat)));
    y=rouhat;
    g=(y*(y/x)^gamma)/(y*(y/x)^gamma+(1-y)*((1-y)/(1-x))^gamma);
    n=n+1;
    urn=unifrnd(0,1,1);
    if urn<=g
       nA=nA+1;
       nrnA=normrnd(muA, sigmaA);
       NRNA=[NRNA nrnA];
       rA=rA+nrnA;
    else
         nB=nB+1;
```

```
            nrnB=normrnd(muB, sigmaB);
            NRNB=[NRNB nrnB];
            rB=rB+nrnB;
        end
    end

    NA=nA;
    NB=nB;
    RHO=NA/N;
    MRHO=[MRHO RHO];
    RA=rA;
    RB=rB;
    TR=RA+RB;
    MTR=[MTR TR];
    h=ttest2(NRNA, NRNB, [], [], 'unequal');
    H=H+h;
end

ERHO=mean(MRHO);
SDRHO=std(MRHO);
ETR=mean(MTR);
VTR=var(MTR);
POWER=H/10000;
y=-ETR-lamda*VTR;

#New design (equal variance)
```

```
function [ETR, VTR, ERHO, SDRHO, POWER, y, newy, ETS, VTS]
=Mnewdesignequalvar(N, muA, muB, sigmaA, sigmaB)
MRHO=[];
MTR=[];
MTS=[];
H=0;
lamda=0.5;
c=0;
for t=1:1:10000
    n=0;
    nA=0;
    nB=0;
    rA=0;
    rB=0;
    NRNA=[];
    NRNB=[];
    sA=0;
    sB=0;
    SIGN=1;

    while SIGN==1
          n=n+1;
          nA=nA+1;
          nrnA=normrnd(muA, sigmaA);
          NRNA=[NRNA nrnA];
          rA=rA+nrnA;
          if nrnA<=c
```

```
        sA=sA+1;
    end
    n=n+1;
    nB=nB+1;
    nrnB=normrnd(muB, sigmaB);
    NRNB=[NRNB nrnB];
    rB=rB+nrnB;
    if nrnB<=c
        sB=sB+1;
    end
    sigmaAhat=std(NRNA)*sqrt((nA-1)/nA);
    sigmaBhat=std(NRNB)*sqrt((nB-1)/nB);
    if sigmaAhat==0 || sigmaBhat==0
        SIGN=1;
    else
        SIGN=0;
    end
    tn=n+1;
end

gamma=2;
e=0.3;

for i=tn:1:N
    x=nA/n;
    muAhat=mean(NRNA);
    muBhat=mean(NRNB);
```

```
sigmaAhat=std(NRNA)*sqrt((nA-1)/nA);
sigmaBhat=std(NRNB)*sqrt((nB-1)/nB);
rouhat=(1-normcdf(c, muBhat, sigmaBhat)
+e*min((1-normcdf(c, muAhat, sigmaAhat)),
(1-normcdf(c, muBhat, sigmaBhat)))
*sign((1-normcdf(c, muBhat, sigmaBhat))
-(1-normcdf(c, muAhat, sigmaAhat))))
/((1-normcdf(c, muAhat, sigmaAhat))
+(1-normcdf(c, muBhat, sigmaBhat)));
y=rouhat;
g=(y*(y/x)^gamma)/(y*(y/x)^gamma+(1-y)*((1-y)/(1-x))^gamma);
n=n+1;
urn=unifrnd(0,1,1);
if urn<=g
   nA=nA+1;
   nrnA=normrnd(muA, sigmaA);
   NRNA=[NRNA nrnA];
   rA=rA+nrnA;
   if nrnA<=c
      sA=sA+1;
   end
else
     nB=nB+1;
     nrnB=normrnd(muB, sigmaB);
     NRNB=[NRNB nrnB];
```

```
            rB=rB+nrnB;
            if nrnB<=c
               sB=sB+1;
            end
        end
    end

    NA=nA;
    NB=nB;
    RHO=NA/N;
    MRHO=[MRHO RHO];
    RA=rA;
    RB=rB;
    TR=RA+RB;
    MTR=[MTR TR];
    SA=sA;
    SB=sB;
    TS=SA+SB;
    MTS=[MTS TS];
    h=ttest2(NRNA, NRNB);
    H=H+h;
end

ERHO=mean(MRHO);
SDRHO=std(MRHO);
ETR=mean(MTR);
VTR=var(MTR);
```

```
ETS=mean(MTS);

VTS=var(MTS);

POWER=H/10000;

y=-ETR-lamda*VTR;

newy=ETS-lamda*VTS;
```

Bibliography

[1] Bai, Z.D., Hu, F. and Zhang L.-X. (2002). Gaussion approximation theorems for urn models and their applications. *Annals of Applied Probability*, 12, 1149-1173.

[2] Bartlett, R.H., Rolloff, D.W., Cornell, R.G., Andrews, A.F., Dillon, P.W. and Zwischenberger, J.B. (1985). Extracorporeal circulation in neonatal repiratory failure: a prospective randomized trial. *Pediatrics*, 76, 479-4877.

[3] Biswas, A. and Mandal, S. (2004). Optimal adaptive designs in phase III clinical trials for continuous responses with covariates. *Advances in Model-Oriented Design and Analysis*, 51-58.

[4] Biswas, A., Bhattacharya, R. and Zhang, L. (2007). Optimal Response-Adaptive Designs for continuous Responses in Phase III Trials. *Biometrical Journal*, 49, 928-940.

[5] Connor E.M., Sperling R.S., Gelber R., Kiselev P., Scott G., et al. (1994). Reduction of maternal-infant transmission of human immunodeficiency virus type 1 with zidovudine treatment. Pediatric AIDS Clinical Trials Group Protocol 076 Study Group. *New England Journal of Medicine*, 331, 1173-1180.

[6] Efron, B. (1971). Forcing a sequential experiment to be balanced. *Biometrika*, 58, 403-417.

[7] Eisele, J.R. (1994). The doubly adaptive biased coin design for sequential clinical trials. *Journal of Statistical Planning and Inference*, 38, 249-261.

[8] Eisele, J.R. and Woodroofe, M.B. (1995). Central limit theorems for doubly adaptive biased coin designs. *Annals of Statistics*, 23, 234-254.

[9] Freedman B. (1987). Equipoise and the ethics of clinical research. *New England Journal of Medicine*, 3, 141-145.

[10] Gallo, P., Chunag-Stein, C., Dragalin, V., Gaydos, B., Krams, M. and Pinheiro, J. (2006). Adaptive designs in clinical drug development-an executive summary of the PhRMA working group. *Journal of Biopharmaceutical Statistics*, 16, 275-283.

[11] Hu, F. and Rosenberger, W. F. (2003). Optimality, variability, Power: evaluating response-adaptive randomization procedures for treatment comparisons. *Journal of American Statistical Association*, 98, 671-678.

[12] Hu, F. and Rosenberger, W. F. (2006). *The theory of response-adaptive randomization in clinical trials*, Wiley Series in Probability and Statistics, Wiley-Interscience, New York.

[13] Hu, F., Rosenberger, W. F. and Zhang, L.-X. (2006). Asymptotically best response-adaptive randomization procedures. *Journal of Statistical Planning and Inference*, 136, 1911-1922.

[14] Hu, F. and Zhang, L.-X. (2004). Asymptotic properties of doubly adaptive biased coin designs for multi-treatment clinical trials. *Annals of Statistics*, 32, 268-301.

[15] Ivanova, A. (2003). A play-the-winner-type urn design with reduced variability. *Metrika*, 58, 1-13.

[16] Ivanova, A., Rosenberger, W.F., Durham, S.D. and Flournoy, N. (2000). A birth and death urn design for randomized clinical trials: Asymptotic methods. *Sankhya B*, 62, 104-118.

[17] Jennison, C. and Turnbull, B. W. (2000). *Group sequential methods with applications to clinical trials*, Boca Raton, Florida: Chapman and Hall/CRC.

[18] Jennison, C. and Turnbull, B.W. (2006). Discussion of "executive summary of the PHRMA working group on adaptive designs in clinical drug development". *Journal of Biopharmaceutical Statistics*, 16, 293-298.

[19] Lellouch, J. and D. Schwartz (1971). Individual ethics and collective ethics. *Review of the International Statistical Institute*, 39, 127-136.

[20] Melfi, V.F., Page, C. and Geraldes, M. (2001). An adaptive randomized design with application to estimation. *Canadian Journal of Statistics*, 29, 107-116.

[21] Pullman, D. and Wang, X. (2001). Adaptive designs, informed consent, and the ethics of research. *Controlled Clinical Trials*, 22, 203-210.

[22] Rosenberger, W. F. (2002). Randomized urn models and sequential design. *Sequential Analysis*, 21, 1-28.

[23] Rosenberger, W. F., Flournoy, N. and Durham, S.D. (1997) Asymptotic normality of maximum likelihood estiamtors from mutiparameter response-driven designs. *Journal of Statistical Planning and Inference*, 60, 69-76.

[24] Rosenberger, W. F. and Hu, F. (2004) Maximizing power and minimizing treatment failures in clinical trials. *Clinical Trials*, 1, 1-7.

[25] Rosenberger, W. F. and Lachin, J. M. (2002). *Randomization in clinical trials: Theory and practice*, Wiley Series in Probability and Statistics, Wiley-Interscience, New York.

[26] Rosenberger, W. F., Stallard, N., Ivanova, A., Harper, C. N. and Ricks, M. L. (2001). Optimal adaptive designs for binary response trials. *Biometrics*, 57, 909-913.

[27] Wei, L. J. (1979). The generalized Polyas urn for sequential medical trials. *Annals of Statistics*, 7, 291-296.

[28] Wei, L. J. and Durham, S. (1978). The randomized play-the-winner rule in medical trials. *Journal of American Statistical Association*, 73, 840-843.

[29] Yao, Q. and Wei, L.J. (1996). Play the winner for phase II/III clinical trials. *Statistics in Medicine*, 15, 2413-2423.

[30] Yi, Y. and Wang, X. (2007). Statistical inference of adaptive designs with binary responses. *Communications in Statistics-Theory and Methods*, 36, 1643-1653.

[31] Yi, Y. and Wang, X. (2008). Asymptotically efficient estimation in response adaptive trials. *Journal of Statistical Planning and Inference*, 138, 2899-2905.

[32] Yi, Y. and Wang, X. (2009). Response adaptive designs with a variance-penalized criterion. *Biometrical Journal*, 51, 763-773.

[33] Zelen, M. (1969). Play the winner rule and the controlled clinical trial. *Journal of American Statistical Association*, 64, 131-146.

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

[35] Zhang, L. and Rosenberger, W. F. (2007). Response-adaptive randomization for survival trials: the parametric approach. *Applied Statistics*, 56, 153-165.