

Abstract

BANERJEE, ANINDITA. Optimal Two-stage designs in Phase-II Clinical Trials. (Under the direction of Dr. Anastasios A. Tsiatis.)

Two-stage designs have been widely used in phase II clinical trials. Such designs are desirable because they allow a decision to be made on whether a treatment is effective or not after the accumulation of the data at the end of each stage. Optimal fixed two-stage designs, where the sample size at each stage is fixed in advance, were proposed by Simon when the primary outcome is a binary response. We propose an adaptive two-stage design which allows the sample size at the second stage to depend on the results at the first stage. Using a Bayesian decision theoretic construct, we derive optimal adaptive two-stage designs. The optimality criterion is to minimize the expected sample size under the null hypothesis value. We further explore optimal adaptive designs that minimize the expected sample size at the alternative hypothesis, at a probability mid-point between the null and alternative hypotheses and a weighted combination of the null, alternative and mid-point value. We also construct an envelope function that gives the lowest expected sample size for any possible value of the response probability. The different designs are compared to Simon's design as well as the envelope function. The designs that minimize the expected sample size at the mid-point between the null and alternative hypotheses and the design that minimizes a weighted average of the response probabilities are closer to the envelope function. Results show that these designs perform better across a range of the response probability values, and generally surpass Simon's design.

OPTIMAL TWO-STAGE DESIGNS IN PHASE-II CLINICAL TRIALS

BY

ANINDITA BANERJEE

A DISSERTATION SUBMITTED TO THE GRADUATE FACULTY OF
NORTH CAROLINA STATE UNIVERSITY

IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

STATISTICS

RALEIGH, NORTH CAROLINA

14TH JULY, 2006

APPROVED BY:

ANASTASIOS A. TSIATIS

CHAIR OF ADVISORY COMMITTEE

MARIE DAVIDIAN

DENNIS BOOS

DAOWEN ZHANG

To my mother Anuradha Banerjee.

Biography

Anindita Banerjee grew up in the lively city of Calcutta in the eastern part of India. She is the only child born to Ambar Nath Banerjee and Anuradha Banerjee. Having completed her Bachelor's degree (honors) in Statistics from Asutosh College, Calcutta University, she joined the M.Stat program at the Indian Statistical Institute (ISI), Calcutta. In 2002, she went on to pursue her doctoral studies at North Carolina State University, Raleigh, North Carolina after earning her Master's degree in statistics from ISI. During this tenure, she worked under the supervision of Dr. Anastasios A. Tsiatis and gained experience as a graduate intern at the Duke Clinical Research Institute, Durham, North Carolina. She will join Pfizer at their New London, Connecticut location as a Statistician Manager after she receives her doctoral degree.

Acknowledgements

I would like to extend my sincere gratitude to my advisor, Dr. Anastasios A. Tsiatis for his encouragement and support throughout my dissertation. I have been fortunate to work under him and have learned a lot from him. Other than being a great statistician, he is also a very humble person. His dedication and perseverance in research have influenced me and helped me to improve my qualities as a researcher. I am grateful to Dr. Marie Davidian, Dr. Dennis Boos and Dr. Daowen Zhang for being on my committee. I am also thankful to Dr. John F. Monahan for his help and support during my program. I would like to thank Terry, Adrian, Angie and Antoinette for the help they provided in the course of my graduate studies. I would take this opportunity to thank my friends and classmates in the department especially Ray, Lovely, Karen, Aarthi, Kirsten, Feng and Yhenneko for extending their help whenever I needed. During my stay at NC State I have been blessed by wonderful friends like Subathra, Jaya, Rinku, Nripa, Ashish, Neelima, Nikhil, Mrin, Shalini and would like to thank them for being with me through thick and thin; lucky to have roommates like Navdip, Nancy, Manpreet, Sridevi and would like to thank them for taking care of a lot of things so that I did not have to worry about them. I would also like to thank my friends at 2303, L5 and 1814 for their wonderful company. Above all, I would like to thank my parents Ambar and Anuradha Banerjee and my aunt Arundhati Mukherjee for their love, sacrifice and moral support. Their faith and confidence in me has helped me to achieve all that I am today. I would also like to thank my fiance Souvik for his constant support and encouragement during my dissertation.

Table of Contents

List of Figures	vii
List of Tables	viii
1 Introduction	1
1.1 Overview of Group sequential designs	1
1.1.1 Group Sequential designs	2
1.1.2 Examples of group sequential designs	3
1.1.3 Common tests for group sequential designs	4
1.2 Simon's two-stage design	6
1.3 Adaptive designs in clinical trials	7
Bibliography	8
2 Adaptive two-stage designs in phase II clinical trials	10
2.1 Introduction	11
2.2 Adaptive Two-stage Designs	13
2.2.1 Assumptions and Notation	13
2.2.2 Methodology	14
2.3 Results	22
2.4 Discussion	25
Bibliography	25
3 Globally optimal two-stage designs for response probabilities	33
3.1 Introduction	35
3.2 Optimal Adaptive Two-stage Designs	37
3.2.1 Assumptions and Notation	37
3.2.2 Weighted Optimal Design	39
3.3 Results	46
3.4 Discussion	57
Bibliography	58

4 Conclusion	65
Bibliography	69

List of Figures

3.1	Comparison graphs for design $\pi_0=0.05$ and $\pi_1=0.25$	48
3.2	Comparison graphs for design $\pi_0=0.1$ and $\pi_1=0.3$	49
3.3	Comparison graphs for design $\pi_0=0.2$ and $\pi_1=0.4$	49
3.4	Comparison graphs for design $\pi_0=0.3$ and $\pi_1=0.5$	50
3.5	Comparison graphs for design $\pi_0=0.4$ and $\pi_1=0.6$	50
3.6	Comparison graphs for design $\pi_0=0.5$ and $\pi_1=0.7$	51
3.7	Comparison graphs for design $\pi_0=0.6$ and $\pi_1=0.8$	51
3.8	Comparison graphs for design $\pi_0=0.7$ and $\pi_1=0.9$	52
3.9	Comparison graphs for design $\pi_0=0.05$ and $\pi_1=0.20$	52
3.10	Comparison graphs for design $\pi_0=0.1$ and $\pi_1=0.25$	53
3.11	Comparison graphs for design $\pi_0=0.2$ and $\pi_1=0.35$	53
3.12	Comparison graphs for design $\pi_0=0.3$ and $\pi_1=0.45$	54
3.13	Comparison graphs for design $\pi_0=0.4$ and $\pi_1=0.55$	54
3.14	Comparison graphs for design $\pi_0=0.5$ and $\pi_1=0.65$	55
3.15	Comparison graphs for design $\pi_0=0.6$ and $\pi_1=0.75$	55
3.16	Comparison graphs for design $\pi_0=0.7$ and $\pi_1=0.85$	56
3.17	Comparison graphs for design $\pi_0=0.8$ and $\pi_1=0.95$	56

List of Tables

2.1	Comparing the expected sample sizes for Simon's design and the adaptive restricted and unrestricted design	26
2.2	Design when $\pi_d=0.2$ and $\alpha = 0.05$	27
2.3	Design when $\alpha = 0.05$ and $\pi_d = 0.20$ contd.	28
2.4	Design when $\alpha = 0.05$ and $\pi_d = 0.15$	29
2.5	Design when $\alpha = 0.05$ and $\pi_d = 0.15$ contd.	30
2.6	Design when $\alpha = 0.05$ and $\pi_d = 0.15$ contd.	31
3.1	Comparing the weighted optimal design with Simon's design when $\pi_d=0.2$ and $\alpha = 0.05$	59
3.2	Comparing the weighted optimal design with Simon's design when $\alpha = 0.05$ and $\pi_d = 0.20$ contd.	60
3.3	Comparing the weighted optimal design with Simon's design when $\alpha = 0.05$ and $\pi_d = 0.15$	61
3.4	Comparing the weighted optimal design with Simon's design when $\alpha = 0.05$ and $\pi_d = 0.15$ contd.	62
3.5	Comparing the weighted optimal design with Simon's design when $\alpha = 0.05$ and $\pi_d = 0.15$ contd.	63

Chapter 1

Introduction

1.1 Overview of Group sequential designs

In a clinical trial or any kind of experiment, data or patients might enter into the trial in a staggered manner, one by one or in groups. The treatment is assigned to the patients when they enter the trial and data is generated when and as we get the responses from the patients. At the beginning of the trial the number of stages and the number of patients to be enrolled in these stages are specified in advance. So at each stage of the trial, on the basis of the data generated till that point of time, we decide either to accept the drug, reject the drug or continue the trial. This design procedure allows us to test the efficacy, toxicity level for a particular drug or choose a better treatment between two treatments (or between placebo and treatment). The former is the kind of testing procedure that we might come across in a phase-II setting, while the latter is usually the case of a large, multi-centered phase-III trial. Due to the ethical, economic and administrative issues involved, sequential designs have become essential for clinical trials. It is imperative that patients always get the best medication, hence we want to assign the best treatment based on the data accumulated at a certain point of time. Also if the drug is not doing well, we would like to stop the trial early, so that we do not assign the inferior treatment to many patients. Again, if we stop the trial early then the

resources can be used to test the next drug in the pipeline. As mentioned above, sequential designs have the liberty to make a decision at the end of every stage so that it suits in the right framework. From an economic point of view, due to interim analysis, identifying a good treatment takes less time and resources and hence is cost efficient. In fact, this is one of the many reasons why sequential designs are preferred to fixed-sample designs. There are other concerns that should be taken into consideration. For example, we should make sure that the design is executed as planned in the protocol, that we are using the correct section of patients, correct treatments, etc. The feature of the design to be able to monitor the data throughout the trial allows us to make a decision about the trial early before there is any problem with the trial. Hence in all these situations sequential designs fit the scenario very well. But many times, during interim analysis, it is difficult to monitor the data continuously one by one, but rather the analysis could be done in groups of individuals or after certain time interval. This gave rise to group sequential designs or multi-stage designs. For example, as stated in Turnbull and Jennison (2000), the Data Monitoring Safety Boards meet periodically to analyze the data after certain time intervals.

1.1.1 Group Sequential designs

The general methodology of a group sequential design is explained with the help of a hypothesis testing problem, where we need to compare two treatments (or a placebo and a treatment). For a K stage sequential design, suppose we have $2m$ patients at every stage, where each of m patients receive each of the two treatments respectively and the data is analyzed after each group of $2m$ responses. At every stage of the trial, $i = 1, 2, \dots, K$ we compute the test statistic and we reject the treatment if the test statistic exceeds a certain critical value. The trial terminates if at any stage we reject the treatment, otherwise it moves on to the next stage. If

at the end of the K th stage, the test statistic falls within the critical region then we accept the drug or else the drug is deemed to be ineffective. The analysis at the end of every stage is based on accumulating data till that point of the trial, which results in different critical values for every stage. These critical values are determined such that they satisfy the type I error constraint and power restriction is satisfied by the group size, m .

1.1.2 Examples of group sequential designs

The efficacy or toxicity of a drug is examined with respect to a standard level of efficacy or toxicity respectively. If the measure of efficacy or toxicity is binary, then the testing problem underlying the sequential design will be a test of binary proportions problem. For example, we might have $H_0 : \pi = \pi_0$ vs. $H_1 : \pi = \pi_1$ where π_0 and π_1 are the proportionality constants with respect to which we are going to perform the test. For instance as mentioned in Simon (1989) we might be interested to check whether the amount of tumor shrinkage is 50% after administering the treatment. In that case the drug would be deemed effective. This is the kind of test we would come across in a typical phase II setting. In case of Phase III trials though, the standard testing procedure is to compare the proportion of two treatments(say A and B) or of a placebo and treatment. So the test parameter could be the difference of proportions ($\pi_A - \pi_B$) or the ratio of proportions ($\frac{\pi_A}{\pi_B}$).

Another obvious example would be continuous responses following a normal distribution with known variances. We mainly test for the mean effect of the treatment in a phase I or phase II setting while we might test the difference of mean responses for two treatments.

1.1.3 Common tests for group sequential designs

Pocock (1977) adapted the idea of a 'repeated significance test' at a constant nominal significance level to analyze accumulating data at a relatively small number of times over the course of a study. Here also we have m groups of patients arriving in the trial in k stages. This design is motivated by the fact that on ethical grounds, fixed sample size designs cannot be implemented when the patients enter into the trial in a sequential manner. Again, sequential designs are not a practical approach in all cases. Due to the fact that sequential designs are inconvenient for certain circumstances, it would be good to check the comparison between the two treatments when the responses are normal in nature when the variance is known. As mentioned earlier, Pocock used the concept of repeated significance testing used by Armitage (1975). The statistical problem is to compare two treatments (or a treatment and a placebo) A and B , when the response is normally distributed with known variance. A group of $2m$ patients in each stage (with a maximum of K stages) is randomized to the two treatment groups with m patients in each group. If we denote X_{A_k} and X_{B_k} as the observed sums from the two treatment groups for the k th stage, the test statistic for Pocock's test $Z_k = \frac{1}{(\sqrt{2}\sigma)mk} (X_{A_k} - X_{B_k})$ follows a standard normal distribution where $X_{A_k} = \sum_{i=1}^m kX_{Ai}$, $X_{B_k} = \sum_{i=1}^m kX_{Bi}$ and $X_{Ai} \sim N(\mu_A, \sigma^2)$, $X_{Bi} \sim N(\mu_B, \sigma^2)$, where σ^2 is known. For any stage $k = 1, \dots, K-1$ if $|Z_k| \geq C_P(k, \alpha)$ we stop and reject H_0 , otherwise we continue to stage $k+1$. At the end of stage K , if $|Z_K| \geq C_P(K, \alpha)$ we stop and reject H_0 , otherwise we stop and accept H_0 . The critical values of $C_P(k, \alpha)$ is chosen such that the test has overall type I error α and is given in Turnbull and Jennison (2000).

In their paper in 1979, O' Brien and Fleming proposed a multiple testing procedure for comparing for two treatments. The method is suggested as an improvement over a fixed one stage trial. Similar to Pocock's design, the treatment is administered to m groups of patients arriving in K stages and data is analyzed periodically when and as responses from a group become available. The hypothesis of no treatment mean difference is given by the test statistic in Turnbull and Jennison (2000) as follows. For stages $k = 1, \dots, K - 1$ if $|Z_k| \geq C_B(k, \alpha) \sqrt{\frac{K}{k}}$ we terminate the trial and reject H_0 , otherwise we continue the trial to the $k + 1$ stage. For the K th stage if $|Z_k| \geq C_B(K, \alpha)$ we terminate the trial and reject H_0 , otherwise we accept H_0 . Z_k is the standardized normal test statistic and $C_B(k, \alpha)$ is chosen such that the overall type-I error is equal to α and different values are given in Turnbull and Jennison (2000). Pocock's test has narrower boundaries initially, allowing greater opportunity for very early stopping, whereas the O' Brien & Fleming test has narrower boundaries later and a smaller maximum sample size. Pocock's test offers greater opportunity to stop early with a low sample size and when it continues to the final analysis it has a higher maximum sample size. The group sequential trial offers reduction in expected sample size over the fixed sample size test when the treatment difference ($|\mu_A - \mu_B|$) is much greater than zero. Pocock has the advantage of lower expected sample size under extreme values of $|\mu_A - \mu_B|$.

Wang and Tsatis (1987) developed a class of optimal tests, such that Pocock and O' Brien & Fleming's testing procedures are special cases. These are two-sided tests indexed by a parameter Δ , which tests the equality of two treatment means when the variance is known. Specifically the test as described in Jennison & Turnbull is given as follows. The assumptions are same as used for the previous two tests. Thus, after stage $k = 1, 2, \dots, K$ we terminate the trial and reject H_0 if $|Z_k| \geq C_{WT}(K, \alpha) \frac{K^{\Delta-1/2}}{k}$ otherwise we continue to the $k + 1$ th stage(or group). At the end of the K th stage we terminate our trial and based on whether $|Z_k| \geq C_{WT}(K, \alpha)$ or

not we reject or accept H_0 respectively. $C_{WT}(K, \alpha)$ is chosen such that the overall level of significance is equal to α and $C_{WT}(K, \alpha)$ is provided for different choices of Δ and K in Wang and Tsiatis (1987). If we allow $\Delta = 0.5$ and $\Delta = 0$ in the above test statistic, we get the Pocock and O' Brien & Fleming test statistic respectively.

The Wang and Tsiatis (1987) test allows the investigator to choose a suitable test (by selecting a particular Δ) depending on whether he wants wider boundaries, smaller expected sample sizes, and so on. Thus group sequential designs are given preference to sequential designs because they allow monitoring of responses of groups of patients rather than monitoring each patient one at a time.

1.2 Simon's two-stage design

We shift our focus to Phase-II trials and will only discuss this phase in the following chapters. It is in this phase where we test the efficacy of the drug after the maximum tolerated dose is determined in Phase I. At the end of this phase, effective drugs move on to further phase III testing while ineffective drugs are screened out from the process, so it is also known as the screening phase. One of the most commonly used group sequential design for this purpose is Simon's fixed two-stage design. The design is based on the testing of hypothesis problem where we test if a particular drug is effective or not. The assumption of this design is given as follows. Patients enter into the study sequentially in two stages and the responses are binary in nature. At the first stage we accrue n_1 patients. If the responses from these patients are less than r_1 then the treatment is declared to be ineffective, otherwise it moves on to the second stage where again a fixed number of patients ($n - n_1$) is treated. If at the end of the second stage the total number of responses are greater than r , the drug is effective. Thus, the parameters of Simon's design

are (n, n_1, r_1, r) . Simon derived optimal and minimax designs for the fixed two-stage group sequential design. In case of the optimal design, for a given null and alternative response probability value, optimal parameter (n, n_1, r_1, r) values are generated such that the design satisfies a given type I and type II error constraints. The definition of optimal being that the expected sample size of this design would be lower than the expected sample size for any other design considered, under the null hypothesis value. For the minimax designs, he derived optimal parameters considering a fixed maximum value of the total sample size, n . The idea is to search the smallest maximum ' n ' such that the two-stage design has a specified type I and type II error.

Although Simon's design is optimal, opportunity exists to derive more efficient designs. This is due to the fact that we could consider flexible designs that would allow the design to be more response oriented. We try to give a brief overview of some of these flexible designs commonly known as adaptive designs.

1.3 Adaptive designs in clinical trials

Adaptive designs allow the trial to be flexible at any stage, throughout the drug-making process. There may be different forms of adaptations implemented in a clinical trial like sample size re-estimation, adaptive allocation, stopping early due to efficacy or futility. These adaptive designs advocate in decreasing sample sizes, decreasing the length of the trial, increasing the probability of assigning the patients to the better treatment and so on. Hence the need for implementing these designs have become manifold, specially in the pharmaceutical industry. For instance Chang, 2005 mentions an adaptive design where we could terminate the trial early if the drug is promising. Though, a cut-off should be determined on the basis of the response rate between the treatment and the control in order to stop

early.

Hommel (2001) describes another adaptive design for a multiple testing problem. This design is flexible to the extent that it allows the hypothesis to modify on the basis of the results of the trial. A lot of careful planning should be done, since the more flexible the trials are, the more they are prone to manipulation. On another note, Tsiatis and Mehta (2003) argue that for testing a treatment difference there always exist a group sequential design that would be more efficient than an adaptive design.

We also generate a new adaptive design but contrary to the above designs it is an optimal adaptive design. This design is based on Simon's fixed two-stage design which has been described earlier. Specifically, based on the results obtained from the first stage the new design changes or adapts the second stage sample size. We derive the optimal adaptive design which has the minimum expected sample size under the null hypothesis value and this has been described in Chapter2, while we consider some more globally optimal designs described elaborately in Chapter3. To summarize and provide some future research ideas we move onto Chapter4.

Bibliography

Armitage, P. (1975). *Sequential Medical Trials*, Blackwell, Oxford.

Hommel, G. (2001). Adaptive modifications of hypotheses after an interim analysis, *Biometrical Journal* **43(5)**: 581–589.

Pocock, S. J. (1977). Group sequential methods in the design and analysis of clinical trials, *Biometrika* **64(2)**: 191–199.

Simon, R. (1989). Optimal two-stage designs for phase ii clinical trials, *Controlled Clinical Trials* **10**: 1–10.

Tsiatis, A. A. and Mehta, C. (2003). On the efficiency of the adaptive design for monitoring clinical trials, *Biometrika* **90(2)**: 367–378.

Turnbull, C. and Jennison, B. W. (2000). *Group Sequential Methods with Applications to Clinical Trials*, Chapman and Hall, London.

Wang, S. K. and Tsiatis, A. A. (1987). Approximately optimal one-parameter boundaries for group sequential trials, *Biometrics* **43(1)**: 193–199.

Chapter 2

Adaptive two-stage designs in phase II clinical trials

Anindita Banerjee and Anastasios A. Tsiatis

ABSTRACT

Two-stage designs have been widely used in phase II clinical trials. Such designs are desirable because they allow a decision to be made on whether a treatment is effective or not after the accumulation of the data at the end of each stage. Optimal fixed two-stage designs, where the sample size at each stage is fixed in advance, were proposed by Simon when the primary outcome is a binary response. This paper proposes an adaptive two-stage design which allows the sample size at the second stage to depend on the results at the first stage. Using a Bayesian decision theoretic construct, we derive optimal adaptive two-stage designs; the optimality criterion being minimum expected sample size under the null hypothesis. Comparisons are made between Simon's two-stage fixed design and the new design with respect to this optimality criterion.

2.1 Introduction

Clinical trials, conducted properly, provide a reliable way of determining whether a drug can be introduced into the market. After the tolerable dose level is determined in a phase I clinical trial, a promising drug moves to phase II testing where the feasibility of the treatment, side effects, toxicity, logistics of administration and cost are evaluated. It is during this phase of testing that investigators want to screen out drugs that are unlikely to be effective as well as move promising drugs to further phase III testing as quickly as possible. Often, the efficacy of a treatment during phase II testing is evaluated with a binary response; i.e. success or failure of treatment, and some minimal acceptable probability of response is set to determine whether the drug is deemed feasible for further testing or not.

Multi-stage or group-sequential designs are especially useful for this purpose. In a multi-stage design, the observed response rate is computed on accumulating data from patients entering the study sequentially at pre-determined stages and, if this observed response rate, at any stage, is sufficiently large or small, the study is terminated and the decision to move to phase III testing or to eliminate this treatment from further consideration is made. Such designs are set up so that a decision must be made before or at some pre-determined final stage and these designs must have with appropriate operating characteristics; i.e. prespecified type I and type II errors. Multi-stage designs are popular because they have been shown to result in the correct decision with the same accuracy as a one-stage design but with smaller average sample sizes.

Although the more stages in a multi-stage design, the better the performance in terms of smaller average sample size, the greatest incremental gains are achieved with fewer stages. That is, the greatest gain in average sample size is seen when moving from a one-stage design to a two-stage design and then less and less as

the number of stages grow. This fact, together with the logistical difficulties involved in setting up multi-stage designs have made two-stage designs especially attractive for phase II clinical trials.

In this paper we are not considering a randomized design which compares the treatment under study to a parallel treatment, but rather, we only consider one treatment group whose probability of response is denoted by π . The statistical problem is given as a hypothesis testing problem. Two values for the probability of response are given in advance, π_0 and π_1 , with $\pi_0 < \pi_1$, where $\pi \leq \pi_0$ represents a drug with low activity which should be screened out for further testing and $\pi \geq \pi_1$ represents a drug with sufficient activity to warrant further phase III testing. Simon (1989) considered two-stage designs. In Simon's design, patients enter in two-stages and the number of patients to be treated in these two-stages are fixed. The procedure is given as follows; there are n_1 patients to be treated in the first stage, if the number of responses from the first stage is less than r_1 then the trial terminates and the drug is declared a failure. Otherwise, it goes on to the next stage where $(n - n_1)$ patients are enrolled. If the response from both stages is greater than r then we declare that the drug shows sufficient efficacy to move to phase III testing, otherwise, it is declared a failure. The type I error is defined as the probability of declaring the drug effective when $\pi \leq \pi_0$ and the type II error is the probability of declaring the drug ineffective when $\pi \geq \pi_1$. Simon found optimal and minimax two-stage designs for the above setup subject to the maximum type I error and type II error being less than or equal to prespecified values of α and β respectively.

We too consider two-stage designs, but, we allow the possibility that the second stage sample size and rejection rule depend on the results of the first stage. Among such adaptive designs, we consider finding the optimal design; i.e. the adaptive

two-stage design that minimizes the average sample size under the null hypothesis $\pi = \pi_0$, subject to the type I and type II error constraints. Such designs are more flexible and since the two-stage designs of Simon are special cases of the adaptive designs, the opportunity exists to obtain more efficient two-stage designs. In section 2, we formulate the problem and give the results in section 3. Discussion of the results are presented in section 4.

2.2 Adaptive Two-stage Designs

2.2.1 Assumptions and Notation

We will assume that patients enter the clinical trial sequentially and the primary response for the i -th patient is a binary response denoted by X_i , where

$$X_i = \begin{cases} 1 & \text{if success ,} \\ 0 & \text{if failure .} \end{cases}$$

After enrolling n patients we observe S_n successes, where $S_n = X_1 + \dots + X_n$ and X_1, \dots, X_n are assumed to be identically and independently distributed with response probability π given by $P(X_i = 1)$. We note that (n, S_n) is a sufficient statistic for π . For our testing problem, we denote by π_0 the probability of response, below which, the experimental drug being tested, will be declared as a low activity drug. Similarly, we define π_1 such that, if the experimental drug has probability of response greater than π_1 , then the drug will be deemed effective. For convenience, and keeping with the usual terminology of hypothesis testing, we denote the null hypothesis by $H_0 : \pi \leq \pi_0$ and the alternative hypothesis by $H_1 : \pi \geq \pi_1$. The sample sizes for the first and second stage are denoted by n_1 and n_2 respectively. The data of the trial are denoted by $R = (R_1, R_2)$, where R_1 denotes the data from the first stage X_1, \dots, X_{n_1} and R_2 denotes the data from the second stage, assuming

the study was not terminated at the first stage, $X_{n_1+1}, \dots, X_{n_1+n_2}$. Being an adaptive two-stage design, the sample size $n_2(R_1)$ at the second stage may be a function of the data observed at the first stage. This includes the possibility that $n_2(R_1) = 0$ which would mean that, for such values of R_1 , the study would be terminated at the first stage and no additional data R_2 would be collected. The decision to reject the null hypothesis or not is denoted by the binary variable D where $D = 1$ corresponds to rejecting the null hypothesis and declaring the drug effective or $D = 0$ where we declare the drug ineffective. The decision D , of course, depends on the data R that are observed from both stages. The prespecified type I and type II errors will be denoted by α and β respectively; thus, designs that we consider must satisfy $P\{D(R) = 1|\pi = \pi_0\} \leq \alpha$ and $P\{D(R) = 0|\pi = \pi_1\} \leq \beta$.

2.2.2 Methodology

A two-stage design, adaptive or not, can be summarized by $\{n_1, n_2(R_1), D(R)\}$; that is, the number of patients n_1 in the first stage, the number of patients $n_2(R_1)$ in the second stage, which could be a function of the data collected at the first stage (including the value 0 if the study is to be terminated at the first stage), and finally, the decision to reject the null hypothesis, or not; i.e. $D(R) = 1$ or 0 which would be a function of all the data collected. Our goal is to find the optimal two-stage design, which we define as the design that minimizes the expected sample size at $\pi = \pi_0$; i.e. $E\{n_1 + n_2(R_1)|\pi = \pi_0\}$, subject to the type I and type II error constraints; $P\{D(R) = 1|\pi\} \leq \alpha$ for $\pi \leq \pi_0$ and $P\{D(R) = 0|\pi\} \leq \beta$ for $\pi \geq \pi_1$. Remark: In the subsequent development we only considered adaptive designs for the simple null hypothesis $\pi = \pi_0$ versus the simple alternative $\pi = \pi_1$. Such tests would also be appropriate for testing the composite null hypothesis $\pi \leq \pi_0$ versus the composite alternative hypothesis $\pi \geq \pi_1$ if the power function was monotone in π . Because of the complexity of the designs that we developed, we were not able to

formally prove this monotonicity property. However, we empirically verified, by numerically computing the power function for every test we developed, that these power curves, without exception, were strictly monotone. Hence, from here on, we will only consider the error constraints for the simple null hypothesis $P\{D(R) = 1|\pi = \pi_0\} \leq \alpha$ and the simple alternative hypothesis $P\{D(R) = 0|\pi = \pi_1\} \leq \beta$.

The problem of finding such an optimal adaptive design satisfying the error constraints can be viewed as a constrained optimization problem which can be solved using Lagrange multipliers. The objective function with the Lagrange multipliers can also be cast as a simple Bayesian decision-theoretic problem using an expected loss function that can be minimized using backward induction. This approach for finding optimal designs was used by Lai (1973) for fully sequential designs and by Barber and Jennison (2002) for group-sequential designs. While Jennison (1987) used a direct search over boundaries to compute optimal group-sequential designs. Toward that end, we define the following loss function

$$L\{n_1, R_1, n_2(R_1), R, D(R), \pi\} = \{n_1 + n_2(R_1)\}I(\pi = \pi_0) + d_0I\{\pi = \pi_0, D(R) = 1\} + d_1I\{\pi = \pi_1, D(R) = 0\}, \quad (2.1)$$

where $I(\cdot)$ is the indicator function and d_0 and d_1 are constants to be determined. We also put prior mass on the parameter π at the two values $\pi = \pi_0$ and $\pi = \pi_1$, say, $P(\pi = \pi_0) = P(\pi = \pi_1) = 1/2$. With this loss function and prior probabilities on π , we derive the expected loss as

$$\begin{aligned} E\{L(\cdot)\} &= E[E\{L(\cdot)|\pi\}] \\ &= .5[E\{n_1 + n_2(R_1)|\pi = \pi_0\} + d_0P\{D(R) = 1|\pi = \pi_0\} + \\ &\quad d_1P\{D(R) = 0|\pi = \pi_1\}]. \end{aligned} \quad (2.2)$$

Since $P\{D(R) = 1|\pi = \pi_0\}$ is the type I error and $P\{D(R) = 0|\pi = \pi_1\}$ is the type

II error, we see that the expected loss given by (2.2), up to a multiplicative constant, is the objective function that would need to be minimized if we wanted to minimize the expected sample size under $\pi = \pi_0$, i.e. $E\{n_1 + n_2(R_1)|\pi = \pi_0\}$, subject to constraints on the type I and type II errors, using Lagrange multipliers d_0 and d_1 . Consequently, the optimization problem now is to find the unconstrained optimal design $\{n_1, n_2(R_1), D(R)\}$ that minimizes (2.2) for any Lagrangian multipliers d_0 and d_1 and then to find the d_0 and d_1 for which the optimal unconstrained design satisfy the constraints; namely, a type I error of α and type II error of β . We also note that the choice of prior mass of .5 at each of the points π_0 and π_1 was arbitrary. Since the values of d_0 and d_1 can take on any values, the objective function (2.2) would, up to a proportionality constant, have been equivalent if we had used any other set of probabilities for the prior mass at π_0 and π_1 .

The strategy we just outlined for constrained optimization using Lagrange multipliers would lead to the desired optimal two-stage sequential design if we could find values d_0 and d_1 which exactly satisfied the type I and type II error constraints. However, because of the discrete nature of binomial probabilities, this cannot be achieved. Instead, we chose d_0 and d_1 that resulted in designs with type I and type II error probabilities that were as close as possible to α and β , respectively, without exceeding them. Consequently, the designs we present may not necessarily be the optimal two-stage sequential design; i.e the design which minimizes the expected sample size at $\pi = \pi_0$ subject to the type I and type II error constraints. However, we would expect such designs to be very close to optimal especially if d_0 and d_1 could be found that resulted in designs with type I and type II errors which were close to the desired α and β . In section 3, we discuss the issue of lower bounds for expected sample sizes in greater detail, demonstrating that the stated strategy will, at least, get us close to the optimal design.

The unconstrained problem; that is, minimizing the expected loss with fixed values

of d_0 and d_1 can be viewed as a Bayesian decision-theoretic problem which can be solved using backward induction as we will now show: Starting backward, the last action to be taken, after all the data R are collected, is the decision $D(R)$ to either reject the null hypothesis, $D(R) = 1$, or not, $D(R) = 0$. Using the law of conditional expectations, the expected loss can be computed as

$$E\{L(\cdot)\} = E[E\{L(\cdot)|R\}],$$

where $E\{L(\cdot)|R\}$ is equal to

$$\begin{aligned} & \{n_1 + n_2(R_1)\}P(\pi = \pi_0|R) \\ & + d_0P(\pi = \pi_0|R)I\{D(R) = 1\} \end{aligned} \tag{2.3}$$

$$+ d_1P(\pi = \pi_1|R)I\{D(R) = 0\}. \tag{2.4}$$

The optimal choice $D(R)$ is the one which minimizes the conditional expectation $E\{L(\cdot)|R\}$, which, by comparing (2.3) to (2.4), is to choose $D(R) = 1$ if

$$d_0P(\pi = \pi_0|R) \leq d_1P(\pi = \pi_1|R),$$

and $D(R) = 0$ otherwise. A simple application of Bayes rule leads us to the decision rule that we should reject H_0 , i.e. $D(R) = 1$, if

$$d_0P(\pi = \pi_0|X_1, \dots, X_{n_1+n_2}) \leq d_1P(\pi = \pi_1|X_1, \dots, X_{n_1+n_2}),$$

or equivalently

$$(1/2)d_0\pi_0^{S_{n_1+n_2}}(1 - \pi_0)^{n_1+n_2-S_{n_1+n_2}} \leq (1/2)d_1\pi_1^{S_{n_1+n_2}}(1 - \pi_1)^{n_1+n_2-S_{n_1+n_2}}.$$

By taking logs, we obtain

$$S_{n_1+n_2} \geq \frac{\log\left(\frac{d_1}{d_0}\right) - (n_1 + n_2) \log \frac{1-\pi_0}{1-\pi_1}}{\log \frac{\pi_0(1-\pi_1)}{\pi_1(1-\pi_0)}}.$$

Hence, $D(R) = 1$ if $S_{n_1+n_2} \geq c_{n_1+n_2}$, and 0 otherwise, where

$$c_{n_1+n_2} = \frac{\log\left(\frac{d_1}{d_0}\right) - (n_1 + n_2) \log \frac{1-\pi_0}{1-\pi_1}}{\log \frac{\pi_0(1-\pi_1)}{\pi_1(1-\pi_0)}}. \quad (2.5)$$

We shall refer to this optimal D by $D_{opt}(R)$. Continuing with the backward induction, the optimal choice of $n_2(R_1)$ is obtained by minimizing the conditional expectation $E[L\{n_1, R_1, n_2(R_1), R, D_{opt}(R)\} | R_1]$ which equals

$$\{n_1 + n_2(R_1)\}P(\pi = \pi_0 | R_1) \quad (2.6)$$

$$+ d_0 P(\pi = \pi_0, S_{n_1+n_2(R_1)} \geq c_{n_1+n_2(R_1)} | R_1) \quad (2.7)$$

$$+ d_1 P(\pi = \pi_1, S_{n_1+n_2(R_1)} < c_{n_1+n_2(R_1)} | R_1). \quad (2.8)$$

Let $b(n, \pi, r)$ and $B(n, \pi, r)$ denote the probability that a binomial random variable with sample size n and probability of success π is equal to r or is greater than or equal to r , respectively, for $r = 0, \dots, n$. Then a simple application of Bayes rule and conditional expectations enables us to compute (2.6)-(2.8) by using the following:

$$P(\pi = \pi_0 | R_1) = \frac{b(n_1, \pi_0, S_{n_1})}{b(n_1, \pi_0, S_{n_1}) + b(n_1, \pi_1, S_{n_1})},$$

$$P(\pi = \pi_0, S_{n_1+n_2} \geq c_{n_1+n_2} | R_1) = B\{n_2, \pi_0, \max(0, c_{n_1+n_2} - S_{n_1})\}P(\pi = \pi_0 | R_1),$$

and

$$\begin{aligned} & P(\pi = \pi_1, S_{n_1+n_2} < c_{n_1+n_2} | R_1) \\ &= [1 - B\{n_2, \pi_1, \max(0, c_{n_1+n_2} - S_{n_1})\}]\{1 - P(\pi = \pi_0 | R_1)\}. \end{aligned} \quad (2.9)$$

We see from the formulae above that the optimal value $n_2(R_1)$ depends on R_1 only through the sufficient statistic (n_1, S_{n_1}) . Thus, to find this optimal value of n_2 , we consider, for each n_1 and $s_1 = 0, \dots, n_1$, the sum of (2.6), (2.7) and (2.8) for different integer values of n_2 starting with n_2 equal to zero until we find the minimum which we denote by $n_{2opt}(n_1, S_{n_1} = s_1)$.

The final step in the backward induction is to find the optimal n_1 ; i.e. the value n_1 that minimizes the unconditional expected loss $E[L\{n_1, R_1, n_{2opt}(n_1, S_{n_1}), R, D_{opt}(R)\}]$, which, computed using $E[E\{L(\cdot) | \pi\}]$, equals

$$\begin{aligned} & .5 \left[n_1 + \sum_{s_1=0}^{n_1} n_{2opt}(n_1, s_1) P(S_{n_1} = s_1 | \pi = \pi_0) \right. \\ & + d_0 \sum_{s_1=0}^{n_1} P(S_{n_1+n_{2opt}(n_1, s_1)} \geq c_{n_1+n_{2opt}(n_1, s_1)} | S_{n_1} = s_1, \pi = \pi_0) P(S_{n_1} = s_1 | \pi = \pi_0) \\ & \left. + d_1 \sum_{s_1=0}^{n_1} P(S_{n_1+n_{2opt}(n_1, s_1)} < c_{n_1+n_{2opt}(n_1, s_1)} | S_{n_1} = s_1, \pi = \pi_1) P(S_{n_1} = s_1 | \pi = \pi_1) \right], \end{aligned}$$

or equivalently,

$$.5 \left[n_1 + \sum_{s_1=0}^{n_1} n_{2opt}(n_1, s_1) b(n_1, \pi_0, s_1) \right] \quad (2.10)$$

$$+ d_0 \sum_{s_1=0}^{n_1} B\{n_{2opt}(n_1, s_1), \pi_0, \max(c_{n_1+n_{2opt}(n_1, s_1)} - s_1, 0)\} b(n_1, \pi_0, s_1) \quad (2.11)$$

$$+ d_1 \sum_{s_1=0}^{n_1} [1 - B\{n_{2opt}(n_1, s_1), \pi_1, \max(c_{n_1+n_{2opt}(n_1, s_1)} - s_1, 0)\}] b(n_1, \pi_1, s_1) \quad (2.12)$$

The optimal n_1 is obtained by evaluating (2.10)+(2.11)+(2.12) for integer values n_1 and finding the value for which this is minimized. This will be referred to as n_{1opt} . Notice that expressions (2.11) and (2.12), when evaluated using n_{1opt} are equal to $d_0\alpha(d_0, d_1)$ and $d_1\beta(d_0, d_1)$, where $\alpha(d_0, d_1)$ and $\beta(d_0, d_1)$ are the type I and type II errors respectively for the unconstrained optimal two-stage design with fixed constants d_0 and d_1 .

To find the optimal design subject to constraints on the error probabilities, we must now search for the values of d_0 and d_1 so that $\alpha(d_0, d_1)$ is as close to the prespecified type I error α without exceeding it and similarly $\beta(d_0, d_1)$ is as close to prespecified type II error β without exceeding it. Because of the discreteness of the binomial distribution, the functions $\alpha(\cdot)$ and $\beta(\cdot)$, as functions of d_0 and d_1 , are discontinuous and piecewise constant and therefore finding solutions to the equations

$$\begin{aligned}\alpha(d_0, d_1) &\approx \alpha \\ \beta(d_0, d_1) &\approx \beta\end{aligned}$$

doesn't lend itself well to algorithms involving gradient searches.

Consequently, we proceed as follows: First we define the objective function $U(d_0, d_1)$ as

$$U(d_0, d_1) = \text{abs}\{\alpha(d_0, d_1) - \alpha\} + \text{abs}\{\beta(d_0, d_1) - \beta\}.$$

Ideally, we would like to find values of d_0 and d_1 where $U(\cdot)$ was identically zero. Because of the discreteness of the problem, this will not occur in general. Therefore, our aim was to narrow our search by finding candidate values for d_0 and d_1 which made $U(\cdot)$ as small as possible and then choose the optimal design; i.e. the design which minimized the expected sample size under H_0 while satisfying

the error constraints, among these candidates. Thus, the numerical problem we faced was to find minimum values of the discontinuous objective function $U(\cdot)$. The hyperplanes generated from the objective functions may divide the parameter space into a large number of regions and it is impractical to examine them all. For such complex surfaces, gradient algorithms are not feasible and instead we used the random search simulated annealing algorithm. The actual algorithm that we implemented is described in detail in Lin and Geyer (1992). We not only successfully found the values of d_0 and d_1 that minimized the objective function, but were also able to identify other values of d_0 and d_1 which were close to the minimum. Although the values of d_0 and d_1 that minimizes $U(d_0, d_1)$ will yield $\alpha(d_0, d_1)$ and $\beta(d_0, d_1)$ which are close to α and β , respectively, they are not necessarily smaller as is required by the constraints. Thus, we had to consider other values of d_0 and d_1 to find the design which gave type I and type II error as close to α and β without exceeding them. Because of the very discrete nature of the objective function, the number of such candidate “close to minimum” values were relatively small, among which, we could identify the desired design. The design identified using this algorithm will be denoted as the optimal Bayes design. As we noted, previously, this is not necessarily the optimal two-stage sequential design.

In some cases, we found that the optimal design resulted in a maximum sample size that was very large. Consequently, we also considered designs where we restricted the maximum sample size to some value n_{max} . This did not create any additional difficulties to our methodology as the only change necessary in our procedure was to restrict the search for $n_{2opt}(n_1, s_1)$ to values less than $n_{max} - n_1$; otherwise, everything else described above remained the same.

2.3 Results

We compared Simon's two-stage design to our optimal two-stage adaptive designs, both with unrestricted and restricted maximum sample size, using the same values of π_0 and π_1 that were considered by Simon (1989) and later published in Piantadosi (1997). For the restricted adaptive design we only considered designs where the maximum sample size did not exceed the corresponding maximum sample size for Simon's design by more than 10%. The results that compare the expected sample sizes under the null hypothesis for these three designs are given in Table 1. Along with the expected sample size we have also presented the type I error (α) and power ($\bar{\beta} = 1 - \beta$) of the test so that we can compare the designs in other aspects. We use the subscript S , AU and AR to denote Simon's designs, the unrestricted adaptive two-stage designs and the restricted adaptive two-stage designs in table 1 respectively. Likewise, in tables 2 and 3 we denote Simon's design and the restricted adaptive two-stage design by S and A respectively. As in Simon (1989), we consider combinations of π_0 and π_1 with difference π_d equal to either .15 or .20, the type I error α equal to .05 and the power $\bar{\beta}$ equal to either 0.90 or 0.80.

We wish to point out that Simon's design only allowed for early stopping at the first stage if the treatment is not effective; that is, only if S_{n_1} were sufficiently small. Since the adaptive designs we consider also allow for early stopping at the first stage if S_{n_1} were sufficiently large or small, this gives our designs an advantage. However, we found that allowing for early stopping, for benefit, has minor effect on the expected sample size and hence comparison is made between the proposed adaptive optimal design and Simon's optimal design.

As is expected by the theory, the smallest expected sample size is always achieved by the unrestricted adaptive design. The restricted adaptive design always has smaller expected sample size than Simon's design as well. This, of course, must be

the case since Simon's design is contained within the class of restricted adaptive designs which are contained within the class of unrestricted adaptive designs. The gains in efficiency were modest. For most cases, the expected sample size under the null hypothesis was 3-5% lower for the adaptive designs as compared to Simon's design and the restricted design often gave results very close to the optimal unrestricted design.

For completeness, we give the design parameters for the optimal restricted adaptive two-stage designs for all the combinations of π_0 , π_1 , α and $\bar{\beta}$ considered above. In Tables 2 and 3, we present the results for $\pi_d = .20$ and $\pi_d = .15$ respectively. Specifically, we give the optimal value for n_1 , the optimal value of n_2 , as a function of S_{n_1} and the value r which defines the rejection region; that is, we reject H_0 when $S_{n_1+n_2} \geq r$. We have also presented the values of parameters for Simon's design, i.e. (n, n_1, n_2, r_1, r) .

To illustrate how to read these tables, consider, for example, the optimal restricted design when $\pi_0 = 0.3$, $\pi_1 = 0.45$ (i.e. $\pi_d = .15$), $\alpha = .05$ and $\beta = .10$ which is found on the right hand side of Table 3a. For this scenario, the first stage sample size, n_1 , is 39. If the number of successes S_{n_1} is less than or equal to 12 or greater than or equal to 21, then the study is stopped at the first stage to reject or not reject the null hypothesis respectively. If S_{n_1} fell between 13 and 20, then an additional n_2 patients would be entered into the study, the value of n_2 depending on S_{n_1} . For example, if $S_{n_1} = 16$, then an additional 75 patients would be entered into the study for a total of $39 + 75 = 114$ patients. If the total number of successes among the 114 patients $S_{n_1+n_2}$ was greater than or equal to $r = 43$, then we would reject the null hypothesis, otherwise not. We also note from this table that if S_{n_1} was between 17 and 20, then the number of additional patients was 81 reflecting the restriction on the maximum sample size.

A counter-intuitive feature of this design is that as S_{n_1} increases, the second stage sample size, n_2 increases till a certain point and then abruptly becomes zero. One of the referees suggested that they believe the reason for this is that, the expected sample size is being minimized under the null hypothesis $H_0 : \pi = \pi_0$ only. Since as S_{n_1} increases, the posterior probability concentrates on π_1 , under which there is no penalty for the sample size. Hence it is not simple to understand where the cut-off for the sample size occurs and is dictated by the Bayesian decision rule for obtaining the optimal adaptive design.

As mentioned earlier, the error rates of the optimal Bayes design are not attained exactly and hence there might be a non-Bayes test with lower expected sample size but higher error rates, which are still less than the specified type I and type II error rates. Since we cannot exclude such possibilities, we have computed lower bounds for the expected sample size, i.e. a value for which the expected sample size of any two-stage design, satisfying the error constraints, must exceed. Denoting the expected sample size of the optimal Bayes design by ess_{opt} and the corresponding type I and type II errors by α_{opt} and β_{opt} , a lower bound can be computed by,

$$ess_{opt} - \{d_0(0.05 - \alpha_{opt}) + d_1(0.2 - \beta_{opt})\}$$

To give an example for the lower bound, consider the optimal unrestricted Bayes design when $\pi_0 = 0.5$, $\pi_1 = 0.7$ (i.e. $\pi_d = 0.20$, $\alpha = 0.05$ and $\beta = 0.2$), which yields an expected sample size 22.95. The corresponding lower bound is 22.81, a difference of 0.14. The lower bounds have been computed for all the cases considered and the maximum difference between the expected sample size for the optimal Bayes design and the lower bound was 0.15. It should be noted that the optimal Bayes design is an optimal design for type I and type II errors that are exactly those achieved by the optimal Bayes design.

All the computations were done using the statistical package R version 2.0.1 (Ihaka and Gentleman, 1996). We would gladly make the software available upon request.

2.4 Discussion

Using a Bayesian decision-theoretic approach for minimizing expected loss through backward induction, we have derived almost optimal two-stage designs that minimize the expected sample size under the null hypothesis subject to constraints on the type I and type II errors. We have shown that such designs are adaptive designs in the sense that the sample size at the second stage and the rejection region depend on the results from the first stage.

We compared the results of the optimal Bayes designs to those derived by Simon who considered fixed sample sizes for the two stages and other adaptive designs which restricted the maximum sample size. We found, as expected, that the optimal Bayes adaptive designs always gave better results than those from Simon's designs. Nevertheless, the gains are modest with a 3-5% decrease in the expected sample size under the null hypothesis. A drawback of the adaptive design is the fact that the second stage sample size is not known in advance which may create difficulty in planning resources for the trial. Whether the modest gains in expected sample size are worth the extra difficulty in conducting and planning for such an adaptive design is not clear. But in any case, we have documented the best one can do with a two-stage design if the criterion is minimizing the expected sample size under the null hypothesis.

Table 2.1: Comparing the expected sample sizes for Simon’s design and the adaptive restricted and unrestricted design

π_0	π_d	α_S	α_{AU}	α_{AR}	β_S	β_{AU}	β_{AR}	ess_S	ess_{AU}	ess_{AR}
0.05	0.20	0.047	0.045	0.046	0.812	0.810	0.802	11.89	10.90	11.03
		0.049	0.045	0.045	0.902	0.902	0.902	16.75	16.67	16.67
	0.15	0.047	0.047	0.047	0.801	0.800	0.800	17.60	17.13	17.45
		0.046	0.049	0.049	0.902	0.900	0.900	26.60	25.83	25.83
	0.10	0.047	0.048	0.049	0.805	0.800	0.809	15.01	14.48	15.01
		0.047	0.050	0.046	0.901	0.903	0.903	22.50	21.98	22.29
0.10	0.20	0.048	0.048	0.048	0.800	0.800	0.802	24.65	23.77	24.65
		0.050	0.046	0.050	0.902	0.901	0.900	36.82	35.43	36.24
	0.15	0.050	0.048	0.050	0.800	0.802	0.802	20.58	20.07	20.18
		0.048	0.050	0.050	0.904	0.900	0.900	30.43	29.21	29.21
	0.20	0.049	0.050	0.046	0.800	0.800	0.801	35.37	34.21	34.77
		0.049	0.050	0.050	0.901	0.900	0.900	51.45	50.23	50.23
0.20	0.20	0.050	0.050	0.050	0.803	0.802	0.801	23.63	23.21	23.21
		0.050	0.049	0.050	0.903	0.901	0.901	34.72	33.68	33.78
	0.15	0.050	0.050	0.050	0.802	0.800	0.801	41.71	40.54	40.80
		0.048	0.050	0.050	0.901	0.900	0.900	60.77	58.46	58.60
	0.40	0.049	0.049	0.050	0.801	0.800	0.801	24.52	24.41	24.43
		0.049	0.050	0.050	0.902	0.900	0.900	35.98	34.95	34.95
0.30	0.20	0.049	0.050	0.050	0.805	0.801	0.801	44.93	43.14	43.32
		0.050	0.050	0.050	0.900	0.900	0.900	63.96	62.08	62.88
	0.15	0.050	0.050	0.050	0.801	0.802	0.801	23.50	22.95	23.08
		0.049	0.050	0.050	0.901	0.901	0.902	34.01	32.88	33.45
	0.40	0.047	0.050	0.050	0.802	0.800	0.800	43.72	42.20	42.20
		0.050	0.050	0.050	0.901	0.900	0.900	62.28	60.64	60.90
0.40	0.20	0.049	0.047	0.047	0.802	0.801	0.801	20.48	20.08	20.08
		0.043	0.050	0.050	0.901	0.901	0.904	29.47	28.15	29.08
	0.15	0.048	0.047	0.049	0.800	0.800	0.803	39.35	38.26	39.01
		0.048	0.050	0.050	0.901	0.901	0.900	55.60	54.06	54.35
	0.70	0.049	0.045	0.045	0.804	0.812	0.812	14.82	14.58	14.58
		0.046	0.050	0.047	0.905	0.905	0.901	21.23	20.75	20.89
0.50	0.20	0.049	0.050	0.048	0.807	0.800	0.800	30.29	29.10	29.62
		0.050	0.050	0.049	0.904	0.900	0.900	43.40	41.59	41.59
	0.80	0.049	0.044	0.049	0.802	0.805	0.802	17.72	17.34	17.56
		0.048	0.048	0.048	0.903	0.904	0.903	24.45	23.88	24.38

Table 2.2: Design when $\pi_d=0.2$ and $\alpha = 0.05$

$\pi_0 = 0.05$ and $\beta = 0.2$							$\pi_0 = 0.05$ and $\beta = 0.1$						
$n_1(A) = 8$ and $n_1(S) = 9$							$n_1(A) = 12$ and $n_1(S) = 9$						
S_{n1}	$n_2(A)$	$n_2(S)$	$n(A)$	$n(S)$	$r(A)$	$r(S)$	S_{n1}	$n_2(A)$	$n_2(S)$	$n(A)$	$n(S)$	$r(A)$	$r(S)$
0	0	0	8	9	2	0	0	0	0	12	9	2	3
1	9	8	17	17	3	2	1	8	21	20	30	3	3
2	10	8	18	17	3	2	2	16	21	28	30	4	3
≥ 3	0	8	8	17	2	2	3	21	21	33	30	5	3
							≥ 4	0	21	12	30	2	3
$\pi_0 = 0.1$ and $\beta = 0.2$							$\pi_0 = 0.1$ and $\beta = 0.1$						
$n_1(A) = 10$ and $n_1(S) = 10$							$n_1(A) = 14$ and $n_1(S) = 18$						
S_{n1}	$n_2(A)$	$n_2(S)$	$n(A)$	$n(S)$	$r(A)$	$r(S)$	S_{n1}	$n_2(A)$	$n_2(S)$	$n(A)$	$n(S)$	$r(A)$	$r(S)$
≤ 1	0	0	10	10	2	1	≤ 1	0	0	14	18	3	2
2-3	19	19	29	29	6	5	2	17	0	31	18	6	6
4	22	19	32	29	7	5	3-5	25	17	39	35	8	6
≥ 5	0	19	10	29	2	5	≥ 6	0	17	14	35	3	6
$\pi_0 = 0.2$ and $\beta = 0.2$							$\pi_0 = 0.2$ and $\beta = 0.1$						
$n_1(A) = 12$ and $n_1(S) = 13$							$n_1(A) = 17$ and $n_1(S) = 19$						
S_{n1}	$n_2(A)$	$n_2(S)$	$n(A)$	$n(S)$	$r(A)$	$r(S)$	S_{n1}	$n_2(A)$	$n_2(S)$	$n(A)$	$n(S)$	$r(A)$	$r(S)$
≤ 2	0	0	12	13	5	3	≤ 3	0	0	17	19	5	4
3	12	0	24	13	8	3	4	19	0	36	19	11	15
4	22	30	34	43	11	12	5	29	35	46	54	14	15
5	32	30	44	43	14	12	6	40	35	57	54	17	15
6-7	35	30	47	43	15	12	7-9	42	35	59	54	18	15
≥ 8	0	30	12	43	5	12	≥ 10	0	35	17	54	5	15
$\pi_0 = 0.3$ and $\beta = 0.2$							$\pi_0 = 0.3$ and $\beta = 0.1$						
$n_1(A) = 16$ and $n_1(S) = 15$							$n_1(A) = 25$ and $n_1(S) = 24$						
S_{n1}	$n_2(A)$	$n_2(S)$	$n(A)$	$n(S)$	$r(A)$	$r(S)$	S_{n1}	$n_2(A)$	$n_2(S)$	$n(A)$	$n(S)$	$r(A)$	$r(S)$
≤ 5	0	0	16	15	8	5	≤ 8	0	0	25	24	10	8
6	14	31	30	46	13	18	9	17	39	42	63	17	24
7	24	31	40	46	17	18	10	29	39	54	63	22	24
8-10	34	31	50	46	21	18	11	37	39	62	63	25	24
≥ 11	0	31	16	46	8	18	12-14	44	39	69	63	28	24
							≥ 15	0	39	25	63	10	24

Table 2.3: Design when $\alpha = 0.05$ and $\pi_d = 0.20$ contd.

$\pi_0 = 0.4$ and $\beta = 0.2$							$\pi_0 = 0.4$ and $\beta = 0.1$						
$n_1(A) = 18$ and $n_1(S) = 16$							$n_1(A) = 26$ and $n_1(S) = 25$						
S_{n1}	$n_2(A)$	$n_2(S)$	$n(A)$	$n(S)$	$r(A)$	$r(S)$	S_{n1}	$n_2(A)$	$n_2(S)$	$n(A)$	$n(S)$	$r(A)$	$r(S)$
≤ 7	0	0	18	16	10	7	≤ 11	0	0	26	25	13	11
8	0	30	18	46	10	23	12	15	41	41	66	21	32
9	18	30	36	48	19	23	13	29	41	55	66	28	32
10	30	30	48	46	25	23	14	37	41	63	66	32	32
11-12	32	30	50	46	26	23	15-17	47	41	73	66	37	32
13	33	30	51	46	27	23	≥ 18	0	41	26	66	13	32
≥ 14	0	30	18	46	10	23							
$\pi_0 = 0.5$ and $\beta = 0.2$							$\pi_0 = 0.5$ and $\beta = 0.1$						
$n_1(A) = 15$ and $n_1(S) = 15$							$n_1(A) = 22$ and $n_1(S) = 24$						
S_{n1}	$n_2(A)$	$n_2(S)$	$n(A)$	$n(S)$	$r(A)$	$r(S)$	S_{n1}	$n_2(A)$	$n_2(S)$	$n(A)$	$n(S)$	$r(A)$	$r(S)$
≤ 8	0	0	15	15	10	8	≤ 11	0	0	22	24	14	13
9	23	28	37	43	24	26	12	12	0	34	24	21	13
10	30	28	45	43	28	26	13	30	0	52	24	32	13
11-13	31	28	46	43	29	26	14	40	37	62	61	38	36
≥ 14	0	28	15	43	10	26	15-18	45	37	67	61	41	36
							≥ 19	0	37	22	61	14	36
$\pi_0 = 0.6$ and $\beta = 0.2$							$\pi_0 = 0.6$ and $\beta = 0.1$						
$n_1(A) = 26$ and $n_1(S) = 27$							$n_1(A) = 22$ and $n_1(S) = 19$						
S_{n1}	$n_2(A)$	$n_2(S)$	$n(A)$	$n(S)$	$r(A)$	$r(S)$	S_{n1}	$n_2(A)$	$n_2(S)$	$n(A)$	$n(S)$	$r(A)$	$r(S)$
≤ 6	0	0	10	11	8	7	≤ 12	0	0	22	19	16	12
7	21	0	31	11	23	7	13-14	0	34	22	53	16	30
8	31	32	41	43	30	30	15	17	34	39	53	28	30
≥ 9	37	32	47	43	34	30	16	28	34	50	53	36	30
							17	34	34	56	53	40	30
							18-19	35	34	57	53	41	30
							≥ 20	0	34	22	53	16	30
$\pi_0 = 0.7$ and $\beta = 0.2$							$\pi_0 = 0.7$ and $\beta = 0.1$						
$n_1(A) = 6$ and $n_1(S) = 6$							$n_1(A) = 13$ and $n_1(S) = 15$						
S_{n1}	$n_2(A)$	$n_2(S)$	$n(A)$	$n(S)$	$r(A)$	$r(S)$	S_{n1}	$n_2(A)$	$n_2(S)$	$n(A)$	$n(S)$	$r(A)$	$r(S)$
≤ 4	0	0	6	6	6	4	≤ 9	0	0	13	15	11	11
5	21	21	27	27	23	22	10	13	0	26	15	22	11
6	22	21	28	27	24	22	11	24	0	37	15	31	11
							≥ 12	27	21	50	36	33	29

Table 2.4: Design when $\alpha = 0.05$ and $\pi_d = 0.15$

$\pi_0 = 0.05$ and $\beta = 0.2$							$\pi_0 = 0.05$ and $\beta = 0.1$						
$n_1(A) = 10$ and $n_1(S) = 10$							$n_1(A) = 20$ and $n_1(S) = 21$						
S_{n1}	$n_2(A)$	$n_2(S)$	$n(A)$	$n(S)$	$r(A)$	$r(S)$	S_{n1}	$n_2(A)$	$n_2(S)$	$n(A)$	$n(S)$	$r(A)$	$r(S)$
0	0	0	10	10	2	0	≤ 1	0	0	20	21	3	1
1	18	19	28	29	4	3	2	22	20	42	41	5	4
2	21	19	31	29	4	3	3-4	23	20	43	41	5	4
3	20	19	30	29	4	3	≥ 5	0	20	20	41	3	4
≥ 4	0	19	10	29	2	3							
$\pi_0 = 0.1$ and $\beta = 0.2$							$\pi_0 = 0.1$ and $\beta = 0.1$						
$n_1(A) = 18$ and $n_1(S) = 18$							$n_1(A) = 19$ and $n_1(S) = 21$						
S_{n1}	$n_2(A)$	$n_2(S)$	$n(A)$	$n(S)$	$r(A)$	$r(S)$	S_{n1}	$n_2(A)$	$n_2(S)$	$n(A)$	$n(S)$	$r(A)$	$r(S)$
≤ 2	0	0	18	18	4	2	≤ 1	0	0	19	21	4	2
3	25	25	43	43	8	7	2	15	0	34	21	6	2
4	26	25	44	43	8	7	3	39	45	58	66	10	10
5	29	25	47	43	9	7	4	51	45	70	66	12	10
≥ 6	0	25	18	43	4	7	5-7	54	45	73	66	13	10
							≥ 8	0	45	19	66	4	10
$\pi_0 = 0.2$ and $\beta = 0.2$							$\pi_0 = 0.2$ and $\beta = 0.1$						
$n_1(A) = 23$ and $n_1(S) = 22$							$n_1(A) = 37$ and $n_1(S) = 37$						
S_{n1}	$n_2(A)$	$n_2(S)$	$n(A)$	$n(S)$	$r(A)$	$r(S)$	S_{n1}	$n_2(A)$	$n_2(S)$	$n(A)$	$n(S)$	$r(A)$	$r(S)$
≤ 5	0	0	23	22	7	5	≤ 8	0	0	37	37	10	8
6	29	50	52	72	15	19	9	28	46	65	83	18	22
7	43	50	66	72	19	19	10	50	46	87	83	24	22
8	54	50	77	72	22	19	11-14	54	46	91	83	25	22
9-10	55	50	78	72	22	19	≥ 15	0	46	37	83	10	22
≥ 11	0	50	23	72	7	19							
$\pi_0 = 0.3$ and $\beta = 0.2$							$\pi_0 = 0.3$ and $\beta = 0.1$						
$n_1(A) = 26$ and $n_1(S) = 27$							$n_1(A) = 39$ and $n_1(S) = 40$						
S_{n1}	$n_2(A)$	$n_2(S)$	$n(A)$	$n(S)$	$r(A)$	$r(S)$	S_{n1}	$n_2(A)$	$n_2(S)$	$n(A)$	$n(S)$	$r(A)$	$r(S)$
≤ 8	0	0	26	27	11	9	≤ 12	0	0	39	40	15	13
9	24	0	50	27	20	9	13	27	0	66	40	25	13
10	40	54	66	81	26	30	14	48	70	87	110	33	40
11	54	54	80	81	31	30	15	64	70	103	110	39	40
12-15	62	54	78	81	34	30	16	75	70	114	110	43	40
≥ 16	0	0	26	81	11	30	17-20	81	70	120	110	45	40
							≥ 21	0	70	39	110	15	40

Table 2.5: Design when $\alpha = 0.05$ and $\pi_d = 0.15$ contd.

$\pi_0 = 0.4$ and $\beta = 0.2$							$\pi_0 = 0.4$ and $\beta = 0.1$						
$n_1(A) = 28$ and $n_1(S) = 26$							$n_1(A) = 45$ and $n_1(S) = 45$						
S_{n1}	$n_2(A)$	$n_2(S)$	$n(A)$	$n(S)$	$r(A)$	$r(S)$	S_{n1}	$n_2(A)$	$n_2(S)$	$n(A)$	$n(S)$	$r(A)$	$r(S)$
≤ 11	0	0	28	26	14	11	≤ 19	0	0	45	45	22	19
12	0	58	28	84	14	40	20	38	59	83	104	40	49
13	37	58	65	84	32	40	21	57	59	102	104	49	49
14	52	58	80	84	39	40	22-25	68	59	113	104	54	49
15	64	58	92	84	45	40	26-27	69	59	114	104	55	49
16-19	65	58	93	84	45	40	≥ 28	0	59	45	104	22	49
≥ 20	0	58	28	84	14	40							
$\pi_0 = 0.5$ and $\beta = 0.2$							$\pi_0 = 0.5$ and $\beta = 0.1$						
$n_1(A) = 25$ and $n_1(S) = 28$							$n_1(A) = 39$ and $n_1(S) = 42$						
S_{n1}	$n_2(A)$	$n_2(S)$	$n(A)$	$n(S)$	$r(A)$	$r(S)$	S_{n1}	$n_2(A)$	$n_2(S)$	$n(A)$	$n(S)$	$r(A)$	$r(S)$
≤ 13	0	0	25	28	16	15	≤ 20	0	0	39	42	23	22
14	37	0	62	28	37	15	21	38	0	77	42	45	22
15	51	0	76	28	45	15	22	54	0	93	42	54	22
16	63	58	88	83	52	48	23	73	63	112	105	65	60
17-18	65	58	90	83	53	48	24-25	76	63	115	105	67	60
19-20	66	58	91	83	54	48	26-29	77	63	116	105	67	60
≥ 21	0	58	25	83	16	48	≥ 30	0	63	39	105	23	60
$\pi_0 = 0.6$ and $\beta = 0.2$							$\pi_0 = 0.6$ and $\beta = 0.1$						
$n_1(A) = 26$ and $n_1(S) = 27$							$n_1(A) = 31$ and $n_1(S) = 34$						
S_{n1}	$n_2(A)$	$n_2(S)$	$n(A)$	$n(S)$	$r(A)$	$r(S)$	S_{n1}	$n_2(A)$	$n_2(S)$	$n(A)$	$n(S)$	$r(A)$	$r(S)$
≤ 16	0	0	26	27	19	17	≤ 19	0	0	31	34	21	21
17	22	0	48	27	34	17	20	48	0	79	34	54	64
18	41	40	67	67	47	46	21	66	0	97	34	66	64
19-23	47	40	73	67	51	46	22	72	61	103	95	70	64
≥ 24	0	40	26	67	19	48	23-27	73	61	104	95	71	64
							≥ 28	0	61	31	95	21	64

Table 2.6: Design when $\alpha = 0.05$ and $\pi_d = 0.15$ contd.

$\pi_0 = 0.7$ and $\beta = 0.2$							$\pi_0 = 0.7$ and $\beta = 0.1$						
$n_1(A) = 19$ and $n_1(S) = 19$							$n_1(A) = 24$ and $n_1(S) = 25$						
S_{n1}	$n_2(A)$	$n_2(S)$	$n(A)$	$n(S)$	$r(A)$	$r(S)$	S_{n1}	$n_2(A)$	$n_2(S)$	$n(A)$	$n(S)$	$r(A)$	$r(S)$
≤ 14	0	0	19	19	16	14	≤ 17	0	0	24	25	19	18
15	31	40	50	59	40	46	18	33	0	57	25	45	18
16-17	45	40	64	59	51	46	19	47	54	71	79	56	61
≥ 18	46	40	65	59	52	46	20-21	61	54	85	79	67	61
							≥ 22	62	54	86	79	68	61
$\pi_0 = 0.8$ and $\beta = 0.2$							$\pi_0 = 0.8$ and $\beta = 0.1$						
$n_1(A) = 9$ and $n_1(S) = 9$							$n_1(A) = 19$ and $n_1(S) = 19$						
S_{n1}	$n_2(A)$	$n_2(S)$	$n(A)$	$n(S)$	$r(A)$	$r(S)$	S_{n1}	$n_2(A)$	$n_2(S)$	$n(A)$	$n(S)$	$r(A)$	$r(S)$
≤ 7	0	0	9	9	9	7	≤ 16	0	0	19	19	17	16
8	19	20	28	29	26	26	17	22	23	41	42	37	37
9	21	20	30	29	28	26	≥ 18	24	23	43	42	39	37

Bibliography

- Barber, S. and Jennison, C. (2002). Optimal asymmetric one-sided group sequential tests, *Biometrika* **89**(1): 49–60.
- Ihaka, R. and Gentleman, R. (1996). R: A language for data analysis and graphics, *Journal of Computational and Graphical Statistics* **5**: 299–314.
- Jennison, C. (1987). Efficient group sequential tests with unpredictable group sizes, *Biometrika* **74**(1): 155–165.
- Lai, T. L. (1973). Optimal stopping and sequential tests which minimize the maximum expected sample size, *Annals of Statistics* **1**: 659–673.
- Lin, D. and Geyer, C. (1992). Computational methods for semiparametric linear regression with censored data, *Journal of Computational and Graphical Statistics* **1**(1): 77–90.
- Piantadosi, S. (1997). *Clinical Trials: A Methodologic Perspective*, Wiley, New York.
- Simon, R. (1989). Optimal two-stage designs for phase ii clinical trials, *Controlled Clinical Trials* **10**: 1–10.

Chapter 3

Globally optimal two-stage designs for response probabilities

Anindita Banerjee and Anastasios A. Tsiatis

ABSTRACT

Phase-II trials provide a platform where, on the basis of the efficacy, ineffective drugs are screened out and promising drugs move on to the next phase. The efficacy of the drug is often evaluated by a binary response, i.e., success or failure of the drug, and is determined by testing the response probability. Two stage designs are widely used for this purpose because they result in correct decisions with the same accuracy as a one-stage trial but with smaller average sample sizes. Simon (1989) proposed optimal fixed two-stage designs which minimize the expected sample size under the null hypothesis. We have derived optimal adaptive designs at the null that perform better than Simon's design, although the gains are modest (Banerjee and Tsiatis, 2005). By adaptive we mean that the second stage sample size will be dependent on the results from the first stage. We further explore optimal adaptive designs that minimize the expected sample size at the alternative hypothesis, at a probability mid-point between the null and alternative hypotheses and a weighted combination of the null, alternative and mid-point value. We also construct an envelope function that gives the lowest expected sample size for any possible value of the response probability. The different designs are compared to each other as well as the envelope function. The designs that minimize the expected sample size at the mid-point between the null and alternative hypotheses and the design

that minimizes a weighted average of the response probabilities are closer to the envelope function. These designs perform better across a range of the parameter values, and generally surpass Simon's design.

3.1 Introduction

Group sequential designs are fairly common in phase-II trials due to the nature of the data. These designs involve inspection of the data at pre-specified stages and a decision to accept or reject a drug or continue the trial can be made at the end of any stage. The more the stages in a group sequential design, the greater the cost involved in setting up the design, but we gain in reducing expected sample sizes as compared to a single stage design. Throughout this chapter we have confined ourselves to two-stage designs only, which is widely used when implementing group sequential designs. We are dealing only with binary responses, i.e., success or failure of a drug, although group sequential designs are not restricted to this type of data. One of the most widely used two-stage designs, used for binary response data, was developed by Simon in 1989. The basis of his design is a testing of hypothesis problem. We test whether a particular treatment is effective or not with the help of the response probability of the drug. Simon considered fixed two-stage designs which are optimal at the null hypothesis.

Constructing a design optimal at the null hypothesis makes the design more conservative, in the sense that it gives more importance to rejecting a bad treatment than accepting a good one. Thus in order to be a little more flexible, we would give allowance to treatments which are likely to be accepted. Hence instead of getting rid of drugs that are inferior or ineffective, we want to give more importance to accepting better and effective treatments. This could be established by considering an optimality criterion where we minimize the expected sample size under the alternative hypothesis in the sense that we are giving more allowance for accepting a better treatment. The optimal designs explained in Banerjee and Tsiatis (2005) focused on designs where the criterion was to get rid of 'bad' treatments as soon as possible, and the optimality function was minimizing the expected sample

size under the null hypothesis. Moreover a design optimal at the null, considers the null hypothesis value only at the expense of response probability values away from the null.

It is interesting to evaluate the behavior of the optimal designs at the different values of the response probability, i.e., to understand how an adaptive design, optimal at the null hypothesis, performs at other values of the response probability. Intuitively, a design optimal at a certain probability value will be best at that value but might not do that well for other values of the response probability.

There is scope for achieving even smaller expected sample sizes by considering newer optimal designs as compared to the optimal design at the null hypothesis value in Banerjee and Tsiatis (2005). We consider designs with different optimality criterion and compare them with Simon's design to see how they fare. In fact, there was only a 3-5% decrease in the expected sample size when we compared our adaptive design optimal at the null hypothesis to Simon's design. A natural thing to do, therefore, would be to consider other optimal adaptive designs which work better than Simon's design in order to determine if we can reduce the expected sample sizes by a greater amount.

In this chapter, we consider optimal designs based on different optimality criteria. For instance, we consider a design, where the optimality criterion is to minimize the expected sample size in-between the null and the alternative hypothesis value. The designs we develop are either for particular values of the response probability or for a combination of the response probabilities. Each design gives us the smallest expected sample size for a certain optimality criterion, but we are interested in developing a design which gives us the lowest expected sample size across the range of the response probability for a particular combination of the null and alternative hypothesis value and therefore we generate an envelope function. We also

develop a number of optimal adaptive designs as mentioned above. Therefore the optimal design that is closest to the envelope function will be the best choice of an optimal adaptive design for a particular choice of the null and the alternative hypothesis value. Pertaining to a certain scenario, an investigator can choose the best adaptive optimal design, which will have the lowest expected sample size for any value of the response probability, in between π_0 and π_1 . In section 3.2 we discuss the adaptive designs derived using different optimality criteria and explain the weighted optimal design in detail. Later, we produce some results in section 3.3 and conclude with discussion in section 3.4.

3.2 Optimal Adaptive Two-stage Designs

3.2.1 Assumptions and Notation

The assumptions for the design are similar as in the previous chapter, which is given as follows. We assume that patients enter the clinical trial sequentially and that the primary response for the i th patient is binary and is denoted by X_i , where

$$X_i = \begin{cases} 1 & \text{if success ,} \\ 0 & \text{if failure .} \end{cases}$$

After enrolling n patients, we observe S_n successes, where $S_n = X_1 + \dots + X_n$ and X_1, \dots, X_n are assumed to be identically and independently distributed with response probability $\pi = P(X_i = 1)$. We note that (n, S_n) is a sufficient statistic for π . For our testing problem, we denote by π_0 the probability of response below which the experimental drug being tested will be declared as a low activity drug. Similarly, we define π_1 such that, if the experimental drug has probability of response greater than π_1 , then the drug will be deemed effective. For convenience, and keeping with the usual terminology of hypothesis testing, we denote the null hypothesis by $H_0 : \pi \leq \pi_0$ and the alternative hypothesis by $H_1 : \pi \geq \pi_1$.

The response probability in-between the null and the alternative value used for constructing other optimal designs will be denoted by π_* . The sample sizes for the first and second stage are denoted by n_1 and n_2 respectively. The data of the trial are denoted by $R = (R_1, R_2)$, where R_1 denotes the data from the first stage X_1, \dots, X_{n_1} and R_2 denotes the data from the second stage, $X_{n_1+1}, \dots, X_{n_1+n_2}$, assuming that the study was not terminated at the first stage. Being an adaptive two-stage design, the sample size $n_2(R_1)$ at the second stage may be a function of the data observed at the first stage. This includes the possibility that $n_2(R_1) = 0$ which would mean that, for such values of R_1 , the study would be terminated at the first stage and no additional data R_2 would be collected. The decision to reject the null hypothesis or not is denoted by the binary variable D , where $D = 1$ corresponds to rejecting the null hypothesis and declaring the drug effective, or $D = 0$, where we declare the drug ineffective. The decision D depends on the data, R , that are observed from both stages. The prespecified type I and type II errors will be denoted by α and β respectively, and so the designs that we consider must satisfy $P\{D(R) = 1 | \pi = \pi_0\} \leq \alpha$ and $P\{D(R) = 0 | \pi = \pi_1\} \leq \beta$.

Banerjee and Tsiatis (2005) describes the procedure by which we can construct optimal designs under the null hypothesis, i.e., minimizing the expected sample size at $\pi = \pi_0$. In this chapter, we construct other optimal designs, i.e., designs that yield lower expected sample sizes at different values of the response probability, π other than the null. All of these designs are adaptive in the sense that the number of patients in the second stage $n_2(R_1)$, is a function of the data in the first stage. We also derive an envelope function which gives us the minimum expected sample sizes at various values of the response probability for a particular combination of π_0 and π_1 . The envelope function acts as a lower bound for the optimal designs such that any other design we derive, depending on the optimality criterion, has expected sample sizes higher than that of the envelope function at that response

probability value. Optimality criteria defined are based on different values of the response probability.

The first optimal design we derive is minimizing the expected sample size under the alternative hypothesis value, π_1 and let this design be defined as optimal(1). The second design, focuses on minimizing the expected sample size for the response probability in between the null and the alternative value, i.e. at $\pi = \pi_*$ ($\pi_0 < \pi_* < \pi_1$), and is denoted by optimal(2). Finally we consider an optimality criteria which is a weighted mixture of different response probabilities denoted by optimal(3). While developing these designs, for convinience a simple null, $\pi = \pi_0$, versus a simple alternative, $\pi = \pi_1$ is tested rather than a composite null hypothesis $\pi \leq \pi_0$ versus a composite alternative hypothesis $\pi \geq \pi_1$. This assumption is valid since we empirically verified that power curves are monotone in nature. Due to the computational complexity in deriving the optimal(3) design, we only dicsuss the procedure in deriving this design.

3.2.2 Weighted Optimal Design

As the name suggests, the weighted optimal design is derived on the basis of the optimality criterion where we minimize the expected sample size under a weighted combination of response probabilities (including the null and the alternative value). This weighted combination could also exclude the null and the alternative hypothesis value, though in this case we have included both of them. The general technique of deriving this optimal design is the same as the method of deriving the optimal design at the null. We first find the optimal unrestricted design that gives us optimal values of $n_1, n_2(R_1)$ and $D(R)$. The next step is to determine the design parameters for which the optimal unconstrained design satisfies the type I and

type II errors constraints. Thus to begin with we define the following loss function

$$L\{n_1, R_1, n_2(R_1), R, D(R), \pi\} = \sum_{i=0}^K I(\pi = \pi_i) \{n_1 + n_2(R_1)\} + d_0^* I\{\pi = \pi_0, D(R) = 1\} + d_1^* I\{\pi = \pi_1, D(R) = 0\}, \quad (3.1)$$

where $\pi_0 < \pi_i < \pi_1$, ($i = 2, \dots, K$), $I(\cdot)$ is the indicator function and d_0^* and d_1^* are constants to be determined. We also put prior mass on the parameter π at the values $\pi = \pi_i$ ($i = 2, \dots, K$), $\pi = \pi_0$ and $\pi = \pi_1$, say, $P(\pi = \pi_i) = p_i$, ($i = 0, \dots, K$). A simple case would be if $K = 2$: the expected sample size would be minimized under π_0 , π_1 and $\pi_i = \pi_*$. The constants d_0^* and d_1^* in (3.1) are interpreted as Lagrange multipliers used in obtaining our optimal design. With this loss function and the prior probabilities on π , we derive the expected loss as

$$\begin{aligned} E\{L(\cdot)\} &= E[E\{L(\cdot)|\pi\}] \\ &= \sum_{i=0}^K p_i E\{n_1 + n_2(R_1)|\pi = \pi_i\} + p_0 d_0^* P\{D(R) = 1|\pi = \pi_0\} \\ &\quad + p_1 d_1^* P\{D(R) = 0|\pi = \pi_1\} \end{aligned} \quad (3.2)$$

or equivalently,

$$\begin{aligned} E\{L(\cdot)\} &= E[E\{L(\cdot)|\pi\}] \\ &= \sum_{i=0}^K p_i E\{n_1 + n_2(R_1)|\pi = \pi_i\} + d_0 P\{D(R) = 1|\pi = \pi_0\} \\ &\quad + d_1 P\{D(R) = 0|\pi = \pi_1\} \end{aligned} \quad (3.3)$$

where the Lagrange multipliers d_0 and d_1 will be equal to $p_0 d_0^*$ and $p_1 d_1^*$ respectively.

$P\{D(R) = 1|\pi = \pi_0\}$ is the type I error and $P\{D(R) = 0|\pi = \pi_1\}$ is the type II error of our design in the expected loss equation given by (3.3). In order to obtain the optimal design under $\pi = \pi_i$, $i = 0, \dots, K$, we minimize $\sum_{i=0}^K w_i E\{n_1 + n_2(R_1)|\pi = \pi_i\}$, (where w_i are the weights and in this case $w_i = p_i$) subject to

constraints on the type I and type II errors, using Lagrange multipliers d_0 and d_1 . Consequently, the optimization problem now is to find the unconstrained optimal design $\{n_1, n_2(R_1), D(R)\}$ that minimizes (3.3) for any Lagrangian multipliers d_0 and d_1 and then to find d_0 and d_1 for which the optimal unconstrained design satisfies the constraints, namely, a type I error of α and type II error of β . We also note that the choice of prior mass at the points π_i ($i = 0, \dots, K$), was arbitrary. Since the values of d_0 and d_1 can take on any values, the objective function (3.3) would, up to a proportionality constant, have been equivalent if we had used any other set of probabilities for the prior mass at π_0 and π_1 .

Using the law of conditional expectations, the expected loss can be computed as

$$E\{L(\cdot)\} = E[E\{L(\cdot)|R\}],$$

where $E\{L(\cdot)|R\}$ is equal to

$$\begin{aligned} & \sum_{i=0}^K \{n_1 + n_2(R_1)\} P(\pi = \pi_i | R) \\ & + d_0 P(\pi = \pi_0 | R) I\{D(R) = 1\} \end{aligned} \tag{3.4}$$

$$+ d_1 P(\pi = \pi_1 | R) I\{D(R) = 0\}. \tag{3.5}$$

The unconstrained optimal design is obtained by implementing the backward induction algorithm as used in Jennison (1987) and Barber and Jennison (2002). We will follow the steps of this algorithm to lead us to the optimal $\{n_1, n_2(R_1), D(R)\}$. Given the complete data and as the name of the method suggests, the obvious step would be to evaluate the optimal decision at the end of the trial. This optimal $D(R)$ is the one which minimizes the conditional expectation $E\{L(\cdot)|R\}$, which, by comparing (3.4) to (3.5), is to choose $D(R) = 1$ if

$$d_0 P(\pi = \pi_0 | R) \leq d_1 P(\pi = \pi_1 | R),$$

and $D(R) = 0$ otherwise. A simple application of Bayes rule leads us to the decision rule that we should reject H_0 , i.e. $D(R) = 1$, if

$$d_0 P(\pi = \pi_0 | X_1, \dots, X_{n_1+n_2}) \leq d_1 P(\pi = \pi_1 | X_1, \dots, X_{n_1+n_2}),$$

or equivalently

$$(1/4)d_0\pi_0^{S_{n_1+n_2}}(1-\pi_0)^{n_1+n_2-S_{n_1+n_2}} \leq (1/4)d_1\pi_1^{S_{n_1+n_2}}(1-\pi_1)^{n_1+n_2-S_{n_1+n_2}}$$

By taking logs, we obtain

$$\begin{aligned} S_{n_1+n_2} &\geq \frac{\log\left(\frac{d_1}{d_0}\right) - (n_1+n_2) \log \frac{1-\pi_0}{1-\pi_1}}{\log \frac{\pi_0(1-\pi_1)}{\pi_1(1-\pi_0)}} \\ &= c_{n_1+n_2} \end{aligned} \tag{3.6}$$

Hence, $D(R) = 1$ if $S_{n_1+n_2} \geq c_{n_1+n_2}$, and 0.

We shall refer to this optimal D by $D_{opt}(R)$. The next step of the backward induction algorithm will be to obtain an optimal value of the sample size at the second stage. At this stage, we have the information from the first stage of the design, so we compute an expected loss conditional on the data from the first stage, R_1 , given as,

$$E\{L(\cdot) | R_1\} = \sum_{i=0}^K \{n_1 + n_2(R_1)\} P(\pi = \pi_i | R_1) \tag{3.7}$$

$$+ d_0 P(\pi = \pi_0, S_{n_1+n_2} \geq c_{n_1+n_2} | R_1) \tag{3.8}$$

$$+ d_1 P(\pi = \pi_1, S_{n_1+n_2} < c_{n_1+n_2} | R_1) \tag{3.9}$$

Let $b(\cdot)$ and $B(\cdot)$ denote the binomial pmf and cdf respectively. Applying the properties of conditional expectation and Bayes rule to (3.8) and (3.9) and using the

facts,

$$P(\pi = \pi_0 | R_1) = \frac{b(n_1, S_{n_1}, \pi_0)p_0}{p_0b(n_1, S_{n_1}, \pi_0) + p_1b(n_1, S_{n_1}, \pi_1) + \sum_{i=0}^K p_i b(n_1, S_{n_1}, \pi_i)} \quad (3.10)$$

$$= prob_1(say)$$

$$P(\pi = \pi_0, S_{n_1+n_2} \geq c_{n_1+n_2} | R_1) = B(n_2, S_{n_1}, \pi_0)P(\pi = \pi_0 | R_1)P(\pi = \pi_0)$$

and

$$P(\pi = \pi_1, S_{n_1+n_2} < c_{n_1+n_2} | R_1) = [1 - B(n_2, S_{n_1}, \pi_1)] P(\pi = \pi_1 | R_1)P(\pi = \pi_1)$$

we compute the expressions (3.8) and (3.9).

The expression (3.7) can be expanded by applying simple Bayes rule,

$$\sum_{i=0}^K P(\pi = \pi_i | R_1) = \frac{\sum_{i=0}^K b(n_1, S_{n_1}, \pi_i)p_i}{b(n_1, S_{n_1}, \pi_0)p_0 + b(n_1, S_{n_1}, \pi_1)p_1 + \sum_{i=0}^K b(n_1, S_{n_1}, \pi_i)p_i} \quad (3.11)$$

$$= prob_2(say)$$

We note that in all the expressions, the second stage sample size depends only on R_1 through the sufficient statistic (n_1, S_{n_1}) . For each value of n_1 and $S_{n_1} = 0, 1, \dots, n_1$ we compute the above expressions and the value for that $n_2(S_{n_1})$ for which $[E\{L(\cdot) | R_1\}]$ is minimum, gives us the optimal value.

After obtaining the optimal decision rule and optimal second stage sample size, the last action would be to compute the optimal n_1 of our design. This is done by minimizing the unconditional loss function, $E[L\{n_1, R_1, n_2(R_1), R, D(R), \pi\}]$ using the fact $E\{L(\cdot)\} = E\{E(L|\pi)\}$. Expanding (3.3) the unconditional loss function

becomes,

$$\sum_{S_{n_1}=0}^{n_1} \{n_1 + n_2(S_{n_1})\} \sum_{i=0}^K P(\pi = \pi_i | S_{n_1}) + \sum_{S_{n_1}=0}^{n_1} P(S_{n_1+n_2} \geq c_{n_1+n_2} | \pi = \pi_0, S_{n_1}) +$$

$$\sum_{S_{n_1}=0}^{n_1} P(S_{n_1+n_2} < c_{n_1+n_2} | \pi = \pi_1, S_{n_1})$$

or equivalently,

$$\sum_{S_{n_1}=0}^{n_1} \{n_1 + n_2(S_{n_1})\} \frac{\sum_{i=0}^K b(n_1, S_{n_1}, \pi_i) p_i}{b(n_1, S_{n_1}, \pi_0) p_0 + b(n_1, S_{n_1}, \pi_1) p_1 + \sum_{i=0}^K b(n_1, S_{n_1}, \pi_i) p_i} +$$

$$\sum_{S_{n_1}=0}^{n_1} B(n_2, \pi_0, \max(c_{n_1+n_2} - S_{n_1}, 0)) b(n_1, \pi_0, S_{n_1}) P(\pi = \pi_0) +$$

$$\sum_{S_{n_1}=0}^{n_1} [1 - B(n_2, \pi_1, \max(c_{n_1+n_2} - S_{n_1}, 0)) b(n_1, \pi_1, S_{n_1}) P(\pi = \pi_1)]$$

or equivalently,

$$\sum_{S_{n_1}=0}^{n_1} \{n_1 + n_2(S_{n_1})\} \frac{\sum_{i=0}^K b(n_1, S_{n_1}, \pi_i) p_i}{b(n_1, S_{n_1}, \pi_0) p_0 + b(n_1, S_{n_1}, \pi_1) p_1 + \sum_{i=0}^K b(n_1, S_{n_1}, \pi_i) p_i} + \quad (3.12)$$

$$p_0 \sum_{S_{n_1}=0}^{n_1} B(n_2, \pi_0, \max(c_{n_1+n_2} - S_{n_1}, 0)) b(n_1, \pi_0, S_{n_1}) + \quad (3.13)$$

$$p_1 \sum_{S_{n_1}=0}^{n_1} [1 - B(n_2, \pi_1, \max(c_{n_1+n_2} - S_{n_1}, 0)) b(n_1, \pi_1, S_{n_1})] \quad (3.14)$$

The expressions (3.13) and (3.14) gives us the type I and type II errors of the design and we denote them by $\alpha(d_0, d_1)$ and $\beta(d_0, d_1)$ respectively, while the expression (3.12) denotes the optimal criteria of our design. Through the above procedure, we

have obtained the optimal unconstrained design. The next step would be to obtain the Lagrange multipliers d_0 and d_1 for which the design satisfies the type I and type II error constraints. Here we are considering the outcome of the drug to be binary in nature. Due to the discrete behavior of the binomial distribution, the optimization techniques should not involve derivatives; which rules out most search algorithms for continuous functions, since they involve gradient search methods. Hence we define a simple objective function,

$$U(d_0, d_1) = \text{abs}\{\alpha(d_0, d_1) - \alpha\} + \text{abs}\{\beta(d_0, d_1) - \beta\}$$

which will help us determine the (d_0, d_1) values. The smaller the difference between the derived $\alpha(d_0, d_1)$ and $\beta(d_0, d_1)$ to the specified α and β respectively, the smaller will be the value of $U(d_0, d_1)$. Our strategy is to minimize this objective function $U(d_0, d_1)$ such that it is almost close to zero and this is done by using a random search algorithm known as simulated annealing, described in detail by Lin and Geyer (1992). The search for the Lagrange multipliers d_0 and d_1 over a three-dimensional surface using a random search pattern. First we generate two numbers randomly from a bivariate normal distribution having mean equal to the initial choice of (d_0, d_1) and a prespecified variance. We derive the type I and type II errors of the design and update the objective function, say $U(\cdot)^{(i)}$. At each run of the parameters (d_0, d_1) we compare the objective functions $U(\cdot)^{(i+1)}$ and $U(\cdot)^{(i)}$. Depending upon the difference of $U(\cdot)^{(i)}$ and $U(\cdot)^{(i+1)}$, say δ , we take the following action. If δ is less than zero, we update the Lagrange multipliers (d_0, d_1) , otherwise we generate a random number from a Bernoulli distribution with some probability which depends on δ and a constant known as the 'cooling factor'. We compare the random variable with δ and if it is less than δ then the (d_0, d_1) values are updated or else we go to the next run. At each run we update the 'cooling factor', which is so chosen such that the function approaches zero as n increases.

The idea is to determine the parameters randomly over a grid of all possible combinations of (d_0, d_1) such that they satisfy the optimal design characteristics. By implementing this technique we were able to identify the optimal design which has minimum expected sample size and satisfied the type I and type II error constraints. But due to the discreteness issue, this design may not be the optimal Bayes design, though it is very close to the optimal design. So, we claim that there cannot be another design that will satisfy the error constraints and have expected sample size smaller than our optimal design and we have computed lower bounds to prove this fact. Infact, the optimal design will be closer to the Bayes optimal design if the computed type I and type II errors $\alpha(d_0, d_1)$ and $\beta(d_0, d_1)$ are closer to the specified α and β respectively without exceeding them.

3.3 Results

Given a range of parameter values of the response probability that we obtained from Simon (1989) and Piantadosi (1997), we generate optimal designs and compare it with one another and Simon's design. For each pair of values of (π_0, π_1) we compute an optimal design at the alternative, at the mid-point between the null and the alternative response probability value and the weighted combination of the response probability at the null, alternative hypothesis and a value in between the null and the alternative. Although, in the latter design we have applied weights of $(1/3, 1/3, 1/3)$ at the three points (π_0, π_1, π_*) ; in general we could consider more than three points. All these designs have type I error rate α equal to 0.05 but we considered two different power values, $\bar{\beta}(= 1 - \beta)$ equal to 0.8 and 0.9. A graphical representation of the comparison of expected sample sizes between the different designs mentioned above is provided along with Simon's design in figures 3.1 through 3.17. In all the graphs, the envelope function has been shown inorder to compare the optimal designs with respect to the envelope function. The

envelope function gives us the lower bound for the expected sample size for all parameter values of the response probability. For a particular choice of π , the envelope function gives us the lowest expected sample size as compared to the other optimal designs developed and so provides a basis for the comparisons. We denote the optimal adaptive design at the null hypothesis value (π_0); alternative hypothesis value (π_1); mid-point (π_*) between the null and the alternative hypothesis value and a weighted combination of π_0 , π_1 and π_* by Null, Alternative, Mid-point and Mixture respectively. The envelope function and Simon's optimal design have also been shown in the graphs.

The typical shape of the curves in these comparison graphs is due to the fact that we are plotting the expected sample sizes against three response probability values: π_0 , π_1 and π_* . Instead if we consider five response probability values the graphs will be smoother, though the trend will remain the same.

The expected sample size plots show that all the designs have typical trends. The adaptive design optimal at the null and Simon's optimal design have an increasing trend. Though, the adaptive design optimal at null beats Simon's design at π_0 , in most cases it does poorly as we move away from the null. Again, the adaptive design optimal at the alternative has a decreasing trend, having the lowest expected sample size at the π_1 value and larger expected sample sizes (as compared to the other designs) at π_0 . The mid-point design is closer to the weighted optimal design in all cases considered. For the mid-point design we apply 1/2 weight at $\pi = \pi_*$ and 1/4 each at π_0 and π_1 . Also for the weighted optimal design we have applied weights on these three response probability values as mentioned earlier. Thus, the plots of these two designs are similar in nature. Rather, if we considered more than three values for the weighted optimal design, the results might get altered. The designs with 90 % power show similar trends as the designs with 80 % power, though the expected sample sizes are larger for β equals 0.1, which is

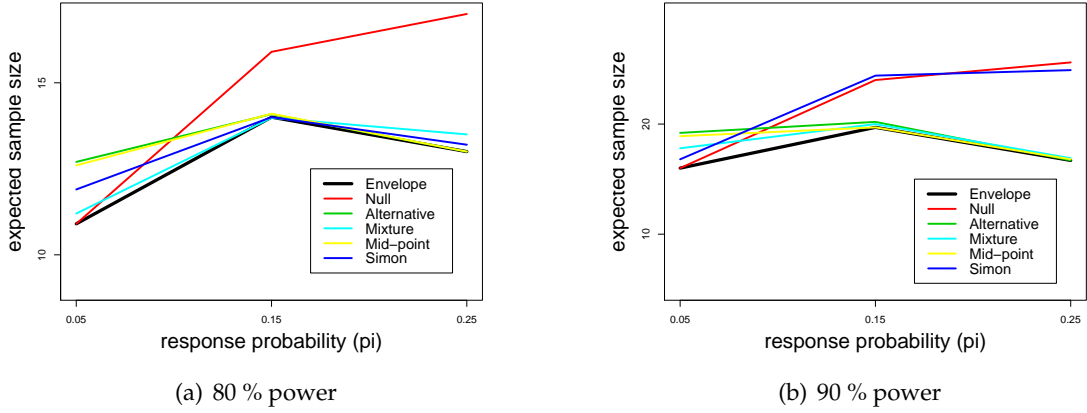
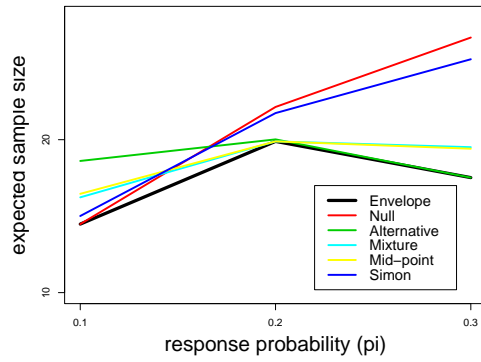


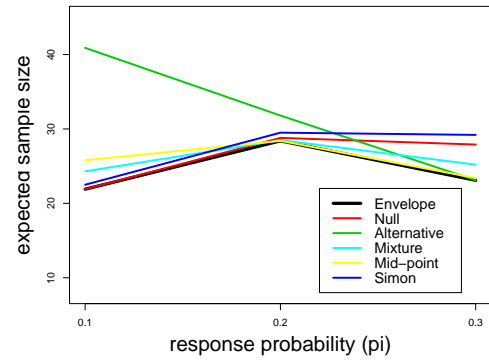
Figure 3.1: Comparison graphs for design $\pi_0=0.05$ and $\pi_1=0.25$

obvious. Due to the discreteness of the binomial distribution the error rates are not exact. Hence while comparing the expected sample size of the adaptive designs, we considered those designs for which the type I error and type II error is less than or equal to 0.05 and 0.2 (or 0.1 as the case may be) respectively. As expected, the weighted optimal design performs fairly well with respect to expected sample size. In most cases it remains close to the envelope function throughout the range of π . This suggests that the weighted optimal adaptive design remains stable for all values of π as compared to other designs. Hence it can be used as an alternative to Simon's design and we have provided a comparison of the actual designs for these two optimal designs in tables 3.1, 3.2, 3.3 and 3.4. For these designs all the computations were done using the statistical package R version 2.0.1 (Ihaka and Gentleman, 1996).

These weighted optimal designs are presented for different combinations of π_0 and $\pi_d (= \pi_1 - \pi_0)$ values. We also provide Simon's optimal design for all these cases so that we can compare it with the weighted optimal design. The subscripts S and W are used to denote Simon's optimal design and the weighted optimal design respectively. Design characteristics for the weighted optimal design are given.

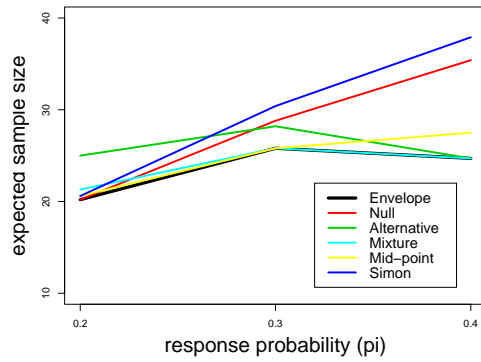


(a) 80 % power

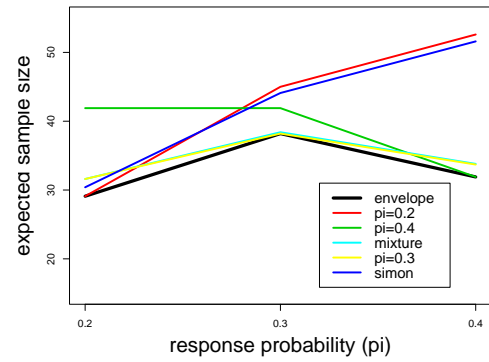


(b) 90 % power

Figure 3.2: Comparison graphs for design $\pi_0=0.1$ and $\pi_1=0.3$

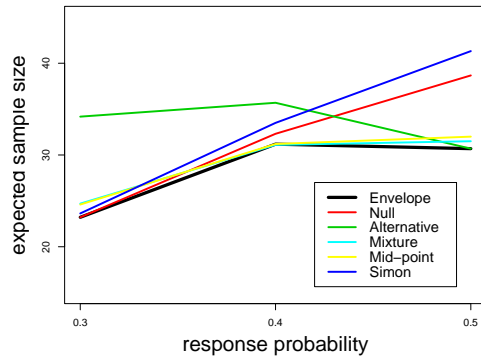


(a) 80 % power

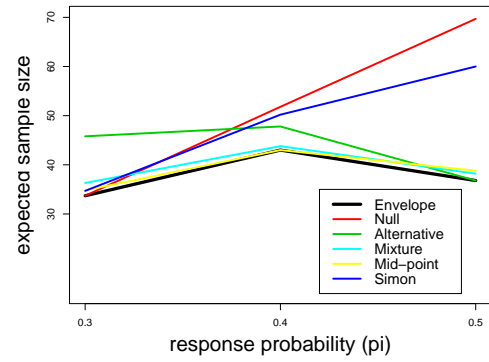


(b) 90 % power

Figure 3.3: Comparison graphs for design $\pi_0=0.2$ and $\pi_1=0.4$

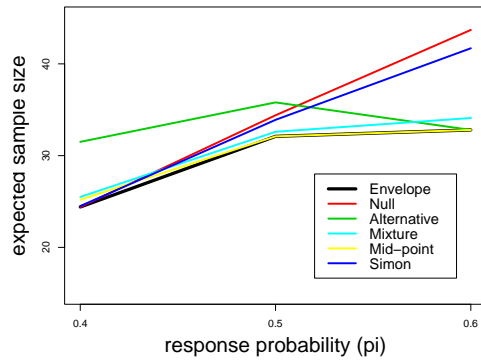


(a) 80 % power

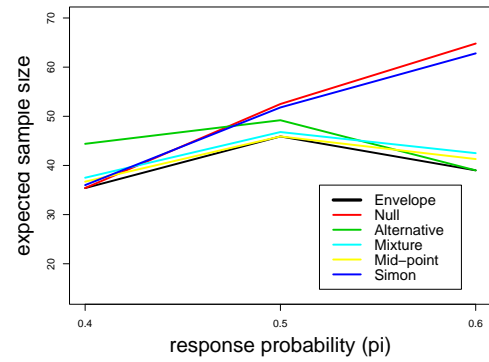


(b) 90 % power

Figure 3.4: Comparison graphs for design $\pi_0=0.3$ and $\pi_1=0.5$

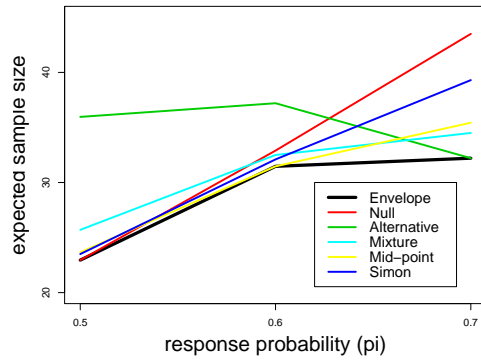


(a) 80 % power

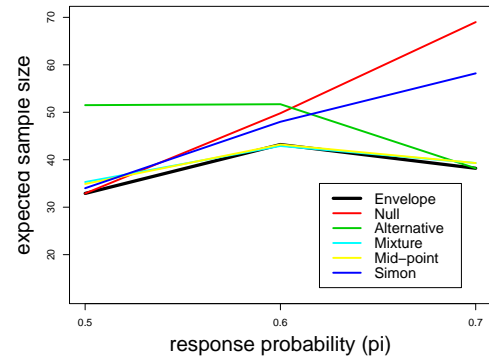


(b) 90 % power

Figure 3.5: Comparison graphs for design $\pi_0=0.4$ and $\pi_1=0.6$

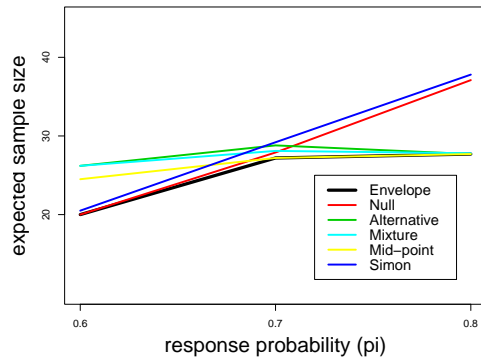


(a) 80 % power

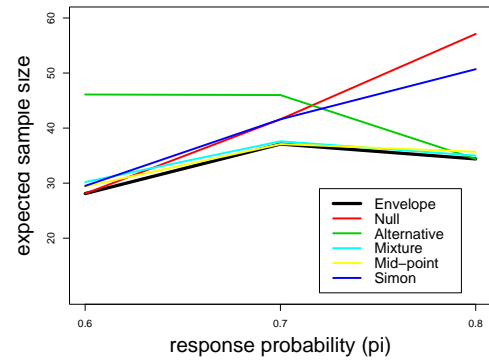


(b) 90 % power

Figure 3.6: Comparison graphs for design $\pi_0=0.5$ and $\pi_1=0.7$

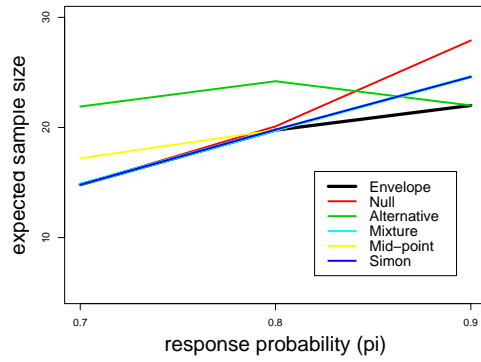


(a) 80 % power

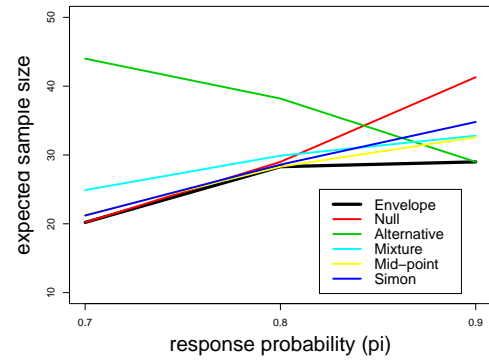


(b) 90 % power

Figure 3.7: Comparison graphs for design $\pi_0=0.6$ and $\pi_1=0.8$

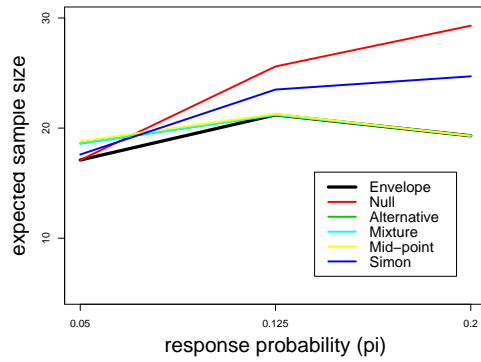


(a) 80 % power

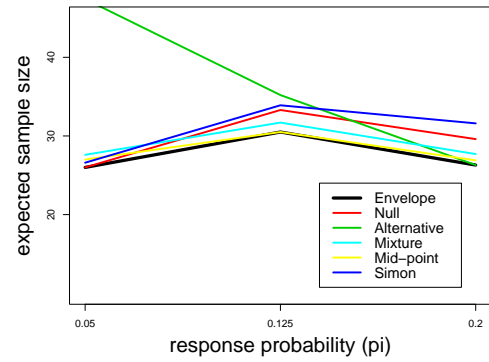


(b) 90 % power

Figure 3.8: Comparison graphs for design $\pi_0=0.7$ and $\pi_1=0.9$

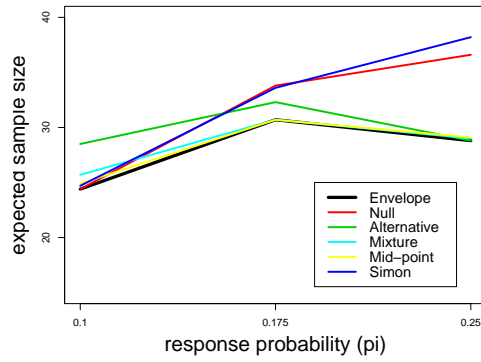


(a) 80 % power

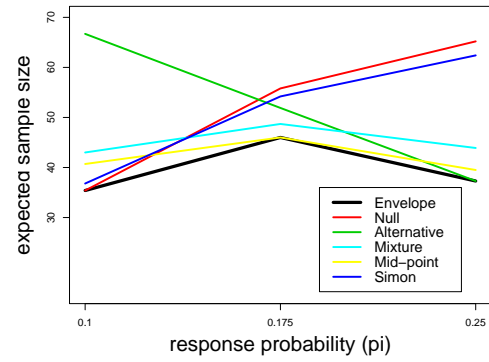


(b) 90 % power

Figure 3.9: Comparison graphs for design $\pi_0=0.05$ and $\pi_1=0.20$

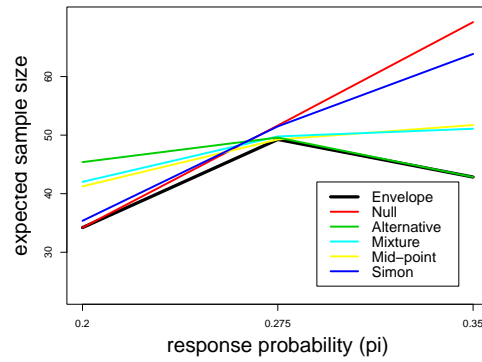


(a) 80 % power

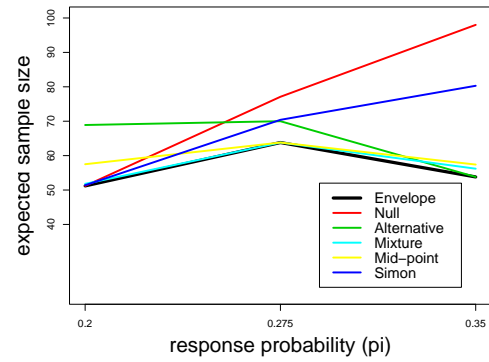


(b) 90 % power

Figure 3.10: Comparison graphs for design $\pi_0=0.1$ and $\pi_1=0.25$

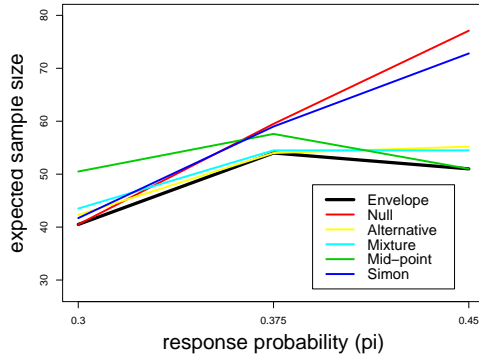


(a) 80 % power

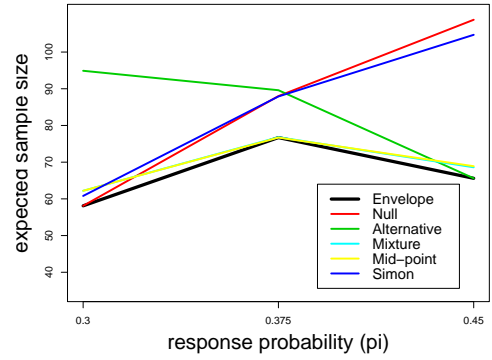


(b) 90 % power

Figure 3.11: Comparison graphs for design $\pi_0=0.2$ and $\pi_1=0.35$

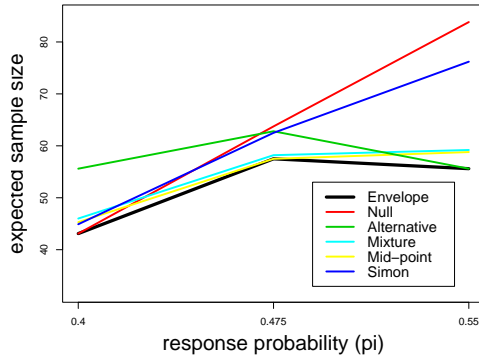


(a) 80 % power

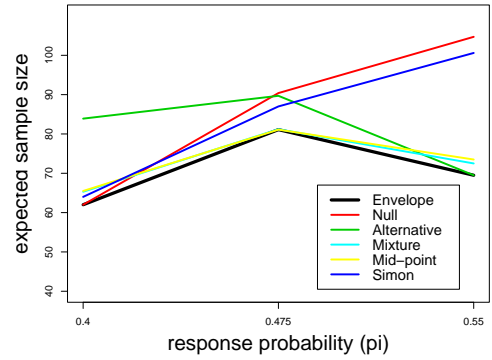


(b) 90 % power

Figure 3.12: Comparison graphs for design $\pi_0=0.3$ and $\pi_1=0.45$

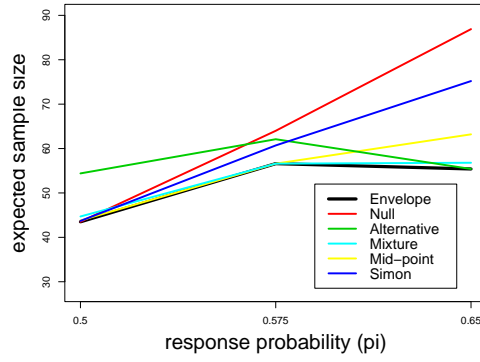


(a) 80 % power

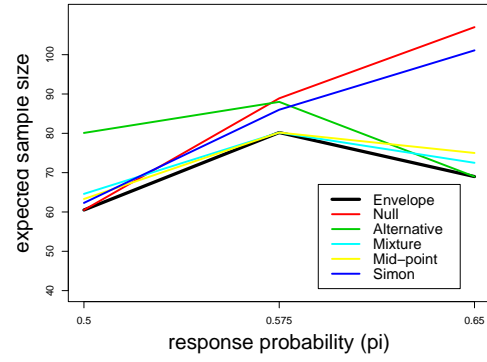


(b) 90 % power

Figure 3.13: Comparison graphs for design $\pi_0=0.4$ and $\pi_1=0.55$

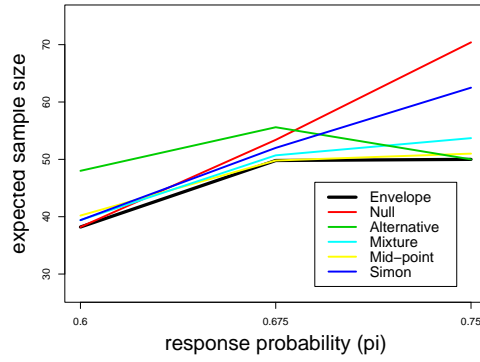


(a) 80 % power

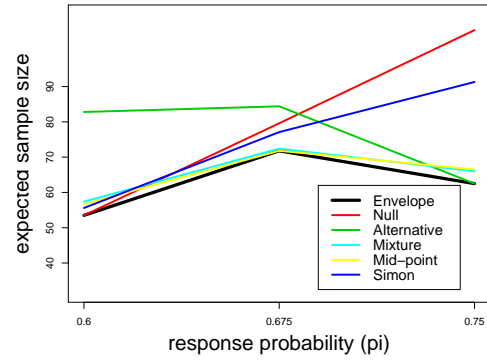


(b) 90 % power

Figure 3.14: Comparison graphs for design $\pi_0=0.5$ and $\pi_1=0.65$

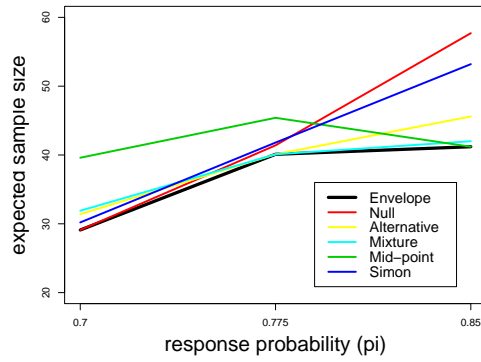


(a) 80 % power

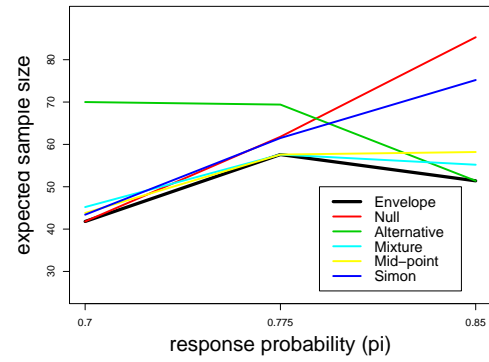


(b) 90 % power

Figure 3.15: Comparison graphs for design $\pi_0=0.6$ and $\pi_1=0.75$

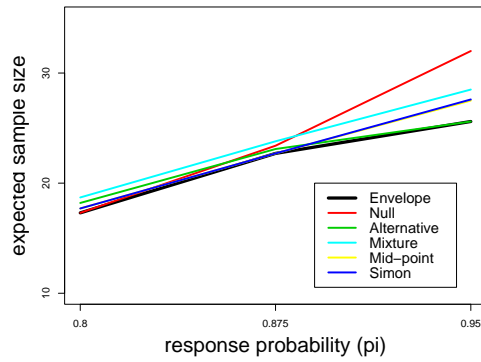


(a) 80 % power

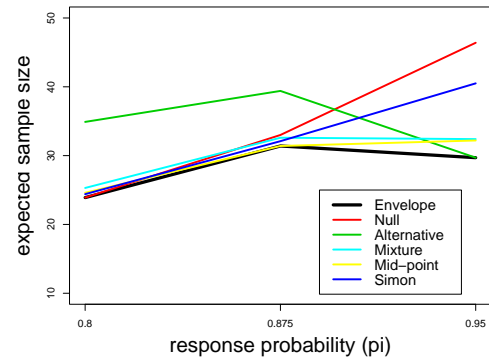


(b) 90 % power

Figure 3.16: Comparison graphs for design $\pi_0=0.7$ and $\pi_1=0.85$



(a) 80 % power



(b) 90 % power

Figure 3.17: Comparison graphs for design $\pi_0=0.8$ and $\pi_1=0.95$

Particularly, the optimal n_1 , the number of successes in the first stage S_{n_1} , the optimal second stage sample size $n_2(W)$, and the rejection rule r . The parameters for Simon's design (n_1, r_1, n, r) are also shown. To illustrate how the weighted optimal design works we consider one example say when $\pi_0 = 0.2$, $\pi_1 = 0.4$ and $\beta = 0.1$ which is on the right hand side of Table 3.1. The adaptive weighted design has an optimal n_1 of 18 while that of Simon's design is 19. For S_{n_1} less than or equal to 3 or greater than 8 the trial stops and we do not need to accrue any more patients in the second stage. Otherwise if the number of successes is between 4 and 7, we have a variable second stage sample size to choose from. If S_{n_1} equals 7 we treat 27 patients in the second stage and a total of 45 patients. Among 45 patients, if the number of responses is greater than or equal to 14 ($=r$) then we declare that the drug is effective. Also the second stage sample size, $n_2(W)$ increases smoothly from zero to a particular value and then decreases smoothly to zero as opposed to the strange nature of the second stage sample size for the design optimal the the null hypothesis value.

3.4 Discussion

Optimal designs for different criteria have been developed. All these designs are adaptive in the sense that the second stage sample size and the rejection rule depend on the results obtained from the first stage. The adaptive design optimal at the null performs well at the null hypothesis value but does poorly with respect to expected sample sizes at other values of the response probability. This is due to the fact that this design emphasizes on the null hypothesis value at the expense of response probability values away from the null. Rather than being conservative, we emphasized on accepting a 'good' treatment and hence constructed the design optimal at the alternative. As expected, this design performed better than Simon's design at π_1 and it is a good design to implement if the criterion is to minimize

the expected sample size under the alternative hypothesis. Some investigators are of the opinion that if the treatment is doing better then we should administer it to as many patients as possible but others are of the opinion of stopping the trial as early as possible. Moreover, the design optimal at the null performs best at π_0 and the design optimal at the alternative performs best at π_1 while they both fail to perform well at other values. We also considered a design optimal at the mid-point between the null and the alternative value. The expected sample size for this design remains close to the weighted optimal for most of the designs we have shown. Again, the expected sample size plot for the weighted optimal design is closest to the envelope function in most of the cases. The reason being that we have given equal emphasis on the null hypothesis value, alternative value and the mid-point value. Therefore, this design can be identified as an optimal design which does uniformly better across the range of the response probability. Rather than giving importance to rejecting a bad drug quickly (choosing the design optimal at null) or giving importance to accepting an effective drug sooner (choosing the design optimal at alternative) we would reduce the bias either way by implementing the weighted optimal design. The weighted optimal design would be a 'reliable' design to use if we were not sure about the behavior of the drug to be tested.

Table 3.1: Comparing the weighted optimal design with Simon's design when $\pi_d=0.2$ and $\alpha = 0.05$

$\pi_0 = 0.05$ and $\beta = 0.2$							$\pi_0 = 0.05$ and $\beta = 0.1$						
$n_1(W) = 12$ and $n_1(S) = 9$							$n_1(W) = 5$ and $n_1(S) = 9$						
S_{n_1}	$n_2(W)$	$n_2(S)$	$n(W)$	$n(S)$	$r(W)$	$r(S)$	S_{n_1}	$n_2(W)$	$n_2(S)$	$n(W)$	$n(S)$	$r(W)$	$r(S)$
0	0	0	12	9	2	0	0	19	0	24	9	4	0
1	0	8	12	17	2	2	1	20	21	25	30	4	3
2	6	8	18	17	3	2	≥ 2	0	21	5	30	2	3
≥ 3	0	8	12	17	2	2							
$\pi_0 = 0.1$ and $\beta = 0.2$							$\pi_0 = 0.1$ and $\beta = 0.1$						
$n_1(W) = 8$ and $n_1(S) = 10$							$n_1(W) = 10$ and $n_1(S) = 18$						
S_{n_1}	$n_2(W)$	$n_2(S)$	$n(W)$	$n(S)$	$r(W)$	$r(S)$	S_{n_1}	$n_2(W)$	$n_2(S)$	$n(W)$	$n(S)$	$r(W)$	$r(S)$
0	0	0	8	10	3	1	0	0	0	10	18	3	2
1	13	0	21	10	5	1	1	19	0	29	18	6	2
2	19	19	27	29	6	5	2	29	0	39	18	8	2
3	13	19	21	29	5	5	3	23	17	33	35	7	6
≥ 4	0	19	8	29	3	5	≥ 4	0	17	10	35	3	6
$\pi_0 = 0.2$ and $\beta = 0.2$							$\pi_0 = 0.2$ and $\beta = 0.1$						
$n_1(W) = 15$ and $n_1(S) = 13$							$n_1(W) = 18$ and $n_1(S) = 19$						
S_{n_1}	$n_2(W)$	$n_2(S)$	$n(W)$	$n(S)$	$r(W)$	$r(S)$	S_{n_1}	$n_2(W)$	$n_2(S)$	$n(W)$	$n(S)$	$r(W)$	$r(S)$
≤ 3	0	0	15	13	6	3	≤ 3	0	0	18	19	6	4
4	18	30	33	43	11	12	4	23	35	41	54	13	15
5	21	30	36	43	12	12	5-6	33	35	51	54	16	15
6	17	30	32	43	11	12	7	27	35	45	54	14	15
≥ 7	0	30	15	43	6	12	≥ 8	0	35	18	54	6	15
$\pi_0 = 0.3$ and $\beta = 0.2$							$\pi_0 = 0.3$ and $\beta = 0.1$						
$n_1(W) = 14$ and $n_1(S) = 15$							$n_1(W) = 26$ and $n_1(S) = 24$						
S_{n_1}	$n_2(W)$	$n_2(S)$	$n(W)$	$n(S)$	$r(W)$	$r(S)$	S_{n_1}	$n_2(W)$	$n_2(S)$	$n(W)$	$n(S)$	$r(W)$	$r(S)$
≤ 4	0	0	14	15	7	5	≤ 8	0	0	26	24	12	8
5	25	0	39	15	17	5	9	24	39	50	63	21	24
6	30	31	44	46	19	18	10-11	34	39	60	63	25	24
7	35	31	49	46	17	18	12	31	39	57	63	24	24
8	20	31	34	46	15	18	≥ 13	0	39	26	63	12	24
≥ 9	0	31	14	46	7	18							

Table 3.2: Comparing the weighted optimal design with Simon's design when $\alpha = 0.05$ and $\pi_d = 0.20$ contd.

$\pi_0 = 0.4$ and $\beta = 0.2$							$\pi_0 = 0.4$ and $\beta = 0.1$						
$n_1(W) = 13$ and $n_1(S) = 16$							$n_1(W) = 23$ and $n_1(S) = 25$						
S_{n_1}	$n_2(W)$	$n_2(S)$	$n(W)$	$n(S)$	$r(W)$	$r(S)$	S_{n_1}	$n_2(W)$	$n_2(S)$	$n(W)$	$n(S)$	$r(W)$	$r(S)$
≤ 5	0	0	13	16	8	7	≤ 9	0	0	23	25	13	11
6	28	0	41	16	22	7	10	27	0	50	25	26	11
7	34	0	47	16	25	7	11	37	0	60	25	31	11
8	30	30	43	46	23	23	12	41	41	64	66	33	32
9	22	30	35	46	19	23	13	37	41	60	66	31	32
≥ 10	0	30	13	16	8	23	14	26	41	49	66	26	32
							≥ 15	0	41	23	66	13	32
$\pi_0 = 0.5$ and $\beta = 0.2$							$\pi_0 = 0.5$ and $\beta = 0.1$						
$n_1(W) = 11$ and $n_1(S) = 15$							$n_1(W) = 28$ and $n_1(S) = 24$						
S_{n_1}	$n_2(W)$	$n_2(S)$	$n(W)$	$n(S)$	$r(W)$	$r(S)$	S_{n_1}	$n_2(W)$	$n_2(S)$	$n(W)$	$n(S)$	$r(W)$	$r(S)$
≤ 5	0	0	11	15	8	8	≤ 13	0	0	28	24	18	13
6	28	0	39	15	25	8	14	0	37	28	61	18	36
7	33	0	44	15	28	8	15	0	37	28	61	18	36
8	31	0	42	15	27	8	16	25	37	53	61	33	36
9	23	28	34	43	22	26	17-18	30	37	58	61	36	36
≥ 10	0	28	11	43	8	26	19	0	37	22	61	14	36
							≥ 20	0	37	28	61	18	36
$\pi_0 = 0.6$ and $\beta = 0.2$							$\pi_0 = 0.6$ and $\beta = 0.1$						
$n_1(W) = 25$ and $n_1(S) = 11$							$n_1(W) = 21$ and $n_1(S) = 19$						
S_{n_1}	$n_2(W)$	$n_2(S)$	$n(W)$	$n(S)$	$r(W)$	$r(S)$	S_{n_1}	$n_2(W)$	$n_2(S)$	$n(W)$	$n(S)$	$r(W)$	$r(S)$
≤ 7	0	0	25	27	8	7	≤ 12	0	0	22	19	16	12
8-17	0	32	25	11	23	7	13-14	0	34	22	53	16	30
18	8	32	41	43	30	30	15	17	34	39	53	28	30
19	12	32	47	43	34	30	16	28	34	50	53	36	30
≥ 20	0	32					17	34	34	56	53	40	30
							18-19	35	34	57	53	41	30
							≥ 20	0	34	22	53	16	30
$\pi_0 = 0.7$ and $\beta = 0.2$							$\pi_0 = 0.7$ and $\beta = 0.1$						
$n_1(W) = 6$ and $n_1(S) = 6$							$n_1(W) = 6$ and $n_1(S) = 15$						
S_{n_1}	$n_2(W)$	$n_2(S)$	$n(W)$	$n(S)$	$r(W)$	$r(S)$	S_{n_1}	$n_2(W)$	$n_2(S)$	$n(W)$	$n(S)$	$r(W)$	$r(S)$
≤ 4	0	0	6	6	6	4	≤ 3	0	0	13	15	11	11
5-6	21	21	27	27	23	22	4	13	0	26	15	22	11
							5	24	0	37	15	31	11
							≥ 12	27	21	50	36	33	29

Table 3.3: Comparing the weighted optimal design with Simon's design when $\alpha = 0.05$ and $\pi_d = 0.15$

$\pi_0 = 0.05$ and $\beta = 0.2$							$\pi_0 = 0.05$ and $\beta = 0.1$						
$n_1(W) = 12$ and $n_1(S) = 10$							$n_1(W) = 18$ and $n_1(S) = 21$						
S_{n_1}	$n_2(W)$	$n_2(S)$	$n(W)$	$n(S)$	$r(W)$	$r(S)$	S_{n_1}	$n_2(W)$	$n_2(S)$	$n(W)$	$n(S)$	$r(W)$	$r(S)$
0	0	0	12	10	2	0	0	0	0	18	21	3	1
1-2	15	19	27	29	4	3	1	13	0	31	21	4	1
≥ 3	0	19	12	29	2	3	2	22	20	40	41	5	4
							3	21	20	39	41	5	4
							≥ 4	0	20	18	21	0	4
$\pi_0 = 0.1$ and $\beta = 0.2$							$\pi_0 = 0.1$ and $\beta = 0.1$						
$n_1(W) = 10$ and $n_1(S) = 18$							$n_1(W) = 14$ and $n_1(S) = 21$						
S_{n_1}	$n_2(W)$	$n_2(S)$	$n(W)$	$n(S)$	$r(W)$	$r(S)$	S_{n_1}	$n_2(W)$	$n_2(S)$	$n(W)$	$n(S)$	$r(W)$	$r(S)$
0	0	0	10	18	3	2	0	0	0	14	21	3	2
1	25	0	35	18	7	7	1	29	0	43	21	8	2
2	36	0	46	18	9	7	2	47	0	61	21	11	2
3	30	25	40	43	8	7	3	47	45	61	66	11	10
≥ 4	0	25	10	43	3	7	4	35	45	49	66	9	10
							≥ 5	0	45	14	66	3	10
$\pi_0 = 0.2$ and $\beta = 0.2$							$\pi_0 = 0.2$ and $\beta = 0.1$						
$n_1(W) = 10$ and $n_1(S) = 22$							$n_1(W) = 14$ and $n_1(S) = 37$						
S_{n_1}	$n_2(W)$	$n_2(S)$	$n(W)$	$n(S)$	$r(W)$	$r(S)$	S_{n_1}	$n_2(W)$	$n_2(S)$	$n(W)$	$n(S)$	$r(W)$	$r(S)$
≤ 1	0	0	10	22	4	5	≤ 1	0	0	14	37	5	8
2	51	0	61	22	18	19	2	56	0	70	37	20	8
3	55	0	65	22	19	19	3	67	0	81	37	23	8
4	51	0	61	22	18	19	4	74	0	88	37	25	8
5	40	0	50	72	15	19	5	70	0	84	37	24	8
≥ 6	0	50	10	72	4	19	6	59	0	73	37	21	8
							7	0	0	14	37	5	8
							8	0	0	14	37	5	8
							≥ 9	0	45	14	66	5	10
$\pi_0 = 0.3$ and $\beta = 0.2$							$\pi_0 = 0.3$ and $\beta = 0.1$						
$n_1(W) = 24$ and $n_1(S) = 27$							$n_1(W) = 40$ and $n_1(S) = 40$						
S_{n_1}	$n_2(W)$	$n_2(S)$	$n(W)$	$n(S)$	$r(W)$	$r(S)$	S_{n_1}	$n_2(W)$	$n_2(S)$	$n(W)$	$n(S)$	$r(W)$	$r(S)$
≤ 7	0	0	24	27	11	9	≤ 12	0	0	40	40	16	13
8	41	0	65	27	26	9	13	45	0	85	40	33	13
9	49	0	73	27	29	9	14	56	70	96	110	37	40
10	54	54	78	81	31	30	15-16	64	70	104	110	40	40
11	46	54	70	81	28	30	17	56	70	96	110	37	40
12	35	54	59	81	24	30	18	40	70	80	110	31	40
≥ 13	0	54	24	81	11	30	≥ 19	0	70	40	110	16	40

Table 3.4: Comparing the weighted optimal design with Simon's design when $\alpha = 0.05$ and $\pi_d = 0.15$ contd.

$\pi_0 = 0.4$ and $\beta = 0.2$							$\pi_0 = 0.4$ and $\beta = 0.1$						
$n_1(A) = 25$ and $n_1(S) = 26$							$n_1(A) = 47$ and $n_1(S) = 45$						
S_{n_1}	$n_2(A)$	$n_2(S)$	$n(A)$	$n(S)$	$r(A)$	$r(S)$	S_{n_1}	$n_2(A)$	$n_2(S)$	$n(A)$	$n(S)$	$r(A)$	$r(S)$
≤ 10	0	0	25	26	14	11	≤ 19	0	0	47	45	24	19
11	49	0	74	26	14	11	20	0	59	47	104	24	49
12	55	58	80	84	40	40	21	62	59	109	104	53	49
13	57	58	82	84	41	40	22	68	59	115	104	56	49
14	53	58	78	84	39	40	23	73	59	120	104	58	49
15	43	58	98	84	34	40	24	64	59	111	104	54	49
≥ 16	0	58	25	84	14	40	25	56	59	103	104	50	49
							≥ 26	0	59	47	104	24	49
$\pi_0 = 0.5$ and $\beta = 0.2$							$\pi_0 = 0.5$ and $\beta = 0.1$						
$n_1(A) = 30$ and $n_1(S) = 28$							$n_1(A) = 43$ and $n_1(S) = 42$						
S_{n_1}	$n_2(A)$	$n_2(S)$	$n(A)$	$n(S)$	$r(A)$	$r(S)$	S_{n_1}	$n_2(A)$	$n_2(S)$	$n(A)$	$n(S)$	$r(A)$	$r(S)$
≤ 15	0	0	30	28	19	15	≤ 22	0	0	43	42	26	22
16	0	58	30	83	19	48	23	52	63	95	105	56	60
17	55	58	85	83	51	48	24	64	63	107	105	63	60
18	57	58	87	83	52	48	25	66	63	109	105	64	60
19	55	58	85	83	51	48	26	64	63	107	105	63	60
20	43	58	73	83	44	48	27	57	63	100	105	59	60
≥ 21	0	58	30	83	19	48	28	43	63	86	105	51	60
							≥ 29	0	63	43	105	26	60
$\pi_0 = 0.6$ and $\beta = 0.2$							$\pi_0 = 0.6$ and $\beta = 0.1$						
$n_1(A) = 24$ and $n_1(S) = 27$							$n_1(A) = 36$ and $n_1(S) = 34$						
S_{n_1}	$n_2(A)$	$n_2(S)$	$n(A)$	$n(S)$	$r(A)$	$r(S)$	S_{n_1}	$n_2(A)$	$n_2(S)$	$n(A)$	$n(S)$	$r(A)$	$r(S)$
≤ 15	0	0	24	27	18	17	≤ 21	0	0	36	34	26	21
16	50	0	74	27	52	17	22	0	61	36	95	26	64
17	53	0	77	27	54	17	23	52	61	88	95	61	64
18	46	40	70	67	49	46	24	64	61	100	95	69	64
19	37	40	24	67	43	46	25	65	61	101	95	70	64
≥ 20	0	40	24	67	18	46	26	61	61	97	95	67	64
							27	49	61	85	95	59	64
							≥ 28	0	61	36	95	26	64

Table 3.5: Comparing the weighted optimal design with Simon's design when $\alpha = 0.05$ and $\pi_d = 0.15$ contd.

$\pi_0 = 0.7$ and $\beta = 0.2$							$\pi_0 = 0.7$ and $\beta = 0.1$						
$n_1(A) = 21$ and $n_1(S) = 19$							$n_1(A) = 23$ and $n_1(S) = 25$						
S_{n_1}	$n_2(A)$	$n_2(S)$	$n(A)$	$n(S)$	$r(A)$	$r(S)$	S_{n_1}	$n_2(A)$	$n_2(S)$	$n(A)$	$n(S)$	$r(A)$	$r(S)$
≤ 14	0	0	21	19	18	14	≤ 16	0	0	23	25	19	18
15	0	40	21	59	18	46	17	51	0	74	25	59	18
16	28	40	49	59	40	46	18	55	0	78	25	62	18
17	37	40	58	59	47	46	19	55	0	78	79	62	61
18	32	40	53	59	43	46	20	42	54	86	79	52	61
19	14	40	35	59	29	46	≥ 21	0	61	23	79	19	61
≥ 20	0	40	21	59	18	46							
$\pi_0 = 0.8$ and $\beta = 0.2$							$\pi_0 = 0.8$ and $\beta = 0.1$						
$n_1(A) = 9$ and $n_1(S) = 9$							$n_1(A) = 17$ and $n_1(S) = 19$						
S_{n_1}	$n_2(A)$	$n_2(S)$	$n(A)$	$n(S)$	$r(A)$	$r(S)$	S_{n_1}	$n_2(A)$	$n_2(S)$	$n(A)$	$n(S)$	$r(A)$	$r(S)$
≤ 7	0	0	9	9	9	7	≤ 14	0	0	17	19	16	16
8-9	20	20	29	29	27	26	15	29	0	46	19	42	16
							16	29	0	46	19	42	16
							≥ 17	0	23	17	42	16	37

Bibliography

- Banerjee, A. and Tsiatis, A. A. (2005). Adaptive designs in phase ii clinical trials, *Statistics in Medicines* **In Press**.
- Barber, S. and Jennison, C. (2002). Optimal asymmetric one-sided group sequential tests, *Biometrika* **89(1)**: 49–60.
- Ihaka, R. and Gentleman, R. (1996). R: A language for data analysis and graphics, *Journal of Computational and Graphical Statistics* **5**: 299–314.
- Jennison, C. (1987). Efficient group sequential tests with unpredictable group sizes, *Biometrika* **74(1)**: 155–165.
- Lin, D. and Geyer, C. (1992). Computational methods for semiparametric linear regression with censored data, *Journal of Computational and Graphical Statistics* **1(1)**: 77–90.
- Piantadosi, S. (1997). *Clinical Trials: A Methodologic Perspective*, Wiley, New York.
- Simon, R. (1989). Optimal two-stage designs for phase ii clinical trials, *Controlled Clinical Trials* **10**: 1–10.

Chapter 4

Conclusion

Two-stage sequential designs have been widely used in phase-II clinical trials. These designs are attractive because they are as accurate as a one-stage design but result in smaller expected sample sizes. They are also more cost effective than setting up multi-stage designs. We have constructed such two-stage designs which are adaptive in nature. Phase II trials are much smaller as compared to phase III trials, so the goal is to accrue less number of patients for testing the efficacy of the drug. Hence, we compute optimal adaptive two-stage designs where the optimality criterion is to minimize the expected sample size under the null hypothesis.

Comparison of this new adaptive design (Banerjee and Tsiatis, 2005) is made with respect to Simon's fixed two-stage design. Simon (1989) proposed a two-stage design where patients enter into the trial sequentially in two phases and the number of patients in these two phases are fixed in advance. Our proposed design allows more flexibility in the sense that the second stage sample size and the rejection rule depends on the results from the first stage. We also assume that the patients enter into the trial sequentially and the response of the i th patient is binary, either success or failure of the drug. Here we only consider one treatment group and test the efficacy of that particular treatment with the help of the response probability of the drug. The approach for obtaining optimal designs is cast as a Bayesian

decision-theoretic problem for minimizing expected loss through backward induction algorithm. We compare the results of optimal Bayes designs and other optimal designs which restricted the maximum sample size to those derived by Simon who considered fixed sample sizes for the two stages and other adaptive designs which restricted the maximum sample size. We found, as expected, that the optimal Bayes adaptive designs always gave better results than Simon's designs. Nevertheless, the gains are modest with a 3-5% decrease in the expected sample size under the null hypothesis.

The designs optimal at the null (π_0) gave more emphasis at the null at the expense of other values away from π_0 , so we explored adaptive designs by considering different optimality criteria (Chapter 3), i.e. minimizing the expected sample size under different values of the response probability. For instance, we use the response probability under the alternative hypothesis as an optimality criterion and another where we use the response probability in between the null and the alternative. We also consider a mixture of the response probabilities to generate the weighted optimal adaptive design. These optimal designs are compared under different scenarios amongst themselves and also with Simon's optimal design. The idea is to be able to identify the 'best' design for a particular pair of null and alternative values of the response probability. By the 'best' design, we mean, the one which has the minimum expected sample size globally across the range of the response probability (π). Hence, we generate an envelope function as a basis for these comparisons. This envelope function provides a lower bound for the expected sample size for any value of the response probability lying between the null (π_0) and the alternative (π_1) probability values. We found that a design optimal at a certain response probability value will have the lowest expected sample size at that value. Moreover, the further we move away from the null hypothesis the more we gain with respect to lowering expected sample sizes. The weighted

optimal design is closest to the envelope function and hence is an adaptive design which is globally optimal across the range of π .

Generally, the efficacy of a drug is tested with the help of the response probability of the drug. Say, for example, the activity of the drug depends on the amount of tumor shrinkage. But for cases, when the tumor is removed, or otherwise, we might be interested in testing the activity of the drug with the help of a time-to-event variable, say one year survival. Two-stage designs have been developed for testing these survival probabilities (Case and Morgan, 2003). As an extension, we are interested in developing optimal adaptive two-stage designs for this statistical testing problem. The optimal criteria would be to minimize the expected sample size and the expected total study length both under the null hypothesis.

Another extension of our research would be to generate confidence intervals for the response probability (π) of the adaptive two-stage designs. Tsiatis et al. (1984) computes the exact confidence interval for the mean when the responses are normal for a group sequential test. The usual method for computing confidence interval will not be appropriate for the adaptive design since the second stage sample size vary depending on the results obtained from the first stage. Thus we will generate exact confidence intervals for adaptive group sequential designs for providing more information on the possible range of values for (π) at the end of the trial.

A drawback of the adaptive design is that the second stage sample size is not known in advance which may create difficulty in planning resources for the trial. Different from the other adaptive designs, our proposed design is optimal with respect to minimizing expected sample sizes. Though the design optimal at π_0 attain modest gains in decreasing expected sample size, it would be the best design one could achieve if the criterion is minimizing the expected sample size under

the null. We also generated a number of adaptive designs having different optimal criterion and a globally optimal design to reduce the bias both at the null and alternative value which allows the investigator to implement the most appropriate design for a particular scenario.

Bibliography

- Banerjee, A. and Tsiatis, A. A. (2005). Adaptive designs in phase ii clinical trials, *Statistics in Medicines* **In Press**.
- Case, L. D. and Morgan, T. M. (2003). Design of phase ii cancer trials evaluating survival probabilities, *BMC Medical Research Methodology* **3(6)**.
"http://bmc.ub.uni-potsdam.de/1471-2288-3-6/1471-2288-3-6.pdf".
- Simon, R. (1989). Optimal two-stage designs for phase ii clinical trials, *Controlled Clinical Trials* **10**: 1–10.
- Tsiatis, A. A., Rosner, G. L. and Mehta, C. R. (1984). Exact confidence intervals following a group sequential test, *Biometrics* **40(3)**: 797–803.