A FRAMEWORK FOR DESIGN OF TWO-STAGE ADAPTIVE PROCEDURES

by

Tatsuki Koyama

B.A., University of California, Berkeley, 1998

M.A., University of Pittsburgh, 2000

Submitted to the Graduate Faculty of
Arts and Sciences in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

University of Pittsburgh

2003

UNIVERSITY OF PITTSBURGH

FACULTY OF ARTS AND SCIENCES

This dissertation was presented

by

Tatsuki Koyama

It was defended on

August 13, 2003

and approved by

Professor Leon J. Gleser (Co-Chairperson)

Professor Allan R. Sampson (Co-Chairperson)

Professor Ori Rosen

Professor H. Samuel Wieand

A FRAMEWORK FOR DESIGN OF TWO-STAGE ADAPTIVE PROCEDURES

Tatsuki Koyama, Ph.D.

University of Pittsburgh, 2003

The main objective of this dissertation is to introduce a framework for two-stage adaptive procedures in which the blind is broken at the end of Stage I. Using our framework, it is possible to control many aspects of an experiment including the Type I error rate, power and maximum total sample size. Our framework also enables us to compare different procedures under the same formulation. We conduct an ANOVA type study to learn the effects of different components of the design specification on the performance characteristics of the resulting design. In addition, we consider conditions for the monotonicity of the power function of a two-stage adaptive procedure.

To foster the practicality of our framework, two extensions are considered. The first one is an application of our framework to the settings with unequal sample sizes. We show how to design a two-stage adaptive procedure having unequal sample sizes for the treatment and control groups. Also we illustrate how to modify an ongoing two-stage adaptive trial when some observations are missing in Stage I and/or in Stage II. Second, we extend the framework to unknown population variance. Our framework can construct a design that incorporates updating the variance estimate at the end of Stage I and modifies the design of Stage II accordingly. All the procedures we present protect the Type I error rate and allow specification of the power and the maximum sample size.

We also consider the problem of switching design objectives between testing noninferiority and testing superiority. Our framework can be used to design a two-stage adaptive procedure that simultaneously tests both noninferiority and superiority hypotheses with controlled error probabilities. The sample size for Stage I is chosen for the main study objective, but that objective may be switched for Stage II based on the unblinded observations from Stage I. Our framework offers a technique to specify certain design criteria such as the various Type I error rates, power and maximum sample sizes.

Preface

This dissertation could not be completed without support of several people. First, I would like to thank my co-advisors, Drs. Leon J. Gleser and Allan R. Sampson. Anyone would be quite fortunate to have just one mentor, but I have two. They have very different personalities, but they are both caring, and they provided me with constant encouragement and guidance to help me grow as a statistician and also as a person. With their bottomless knowledge and keen insight, they navigated this research in a valuable direction. I would also like to thank my committee members, Drs. Ori Rosen and H. Samuel Wieand for their constructive comments which improved this dissertation.

Other professors in the Department of Statistics, Drs. Satish Iyengar, Tom Savits and Henry Block were all helpful and willing to answer my questions whenever I needed their expertise.

I have made several good friends in the five years since I came to Pittsburgh. In early days of graduate work, Jon Getz, Eshetu Wondmagegnehu, Hakan Gogtas and I spent much time together working on homework and on class projects. Coming from four different corners of the world, we formed a very tight group here in Pittsburgh. I will always cherish our friendship.

I would also like to thank Katherine McDonald for always being such a good friend. When my life as a graduate student seems to get a little bit too tense, I always find relief in casual conversations with Katherine. She has also answered my countless questions about English (Is this supposed to be “a” or “the” ?), and helped me to write better. Thanks go to Melissa Ziegler who also helped me with English and proofread some of the manuscripts.

I had such good relationships with everybody in the Department of Statistics, professors, friends and staff members. My sincere gratitude goes to them. Regrettably, I cannot list them all here, but I shall not forget the support I have received from all of them.

I would like to thank my parents, Tamiki & Chikako Koyama. Although they live far across the Pacific, they have always provided me with unconditional support.

And finally, I would like to give my sincerest thanks to Mari. As a friend, fiancée and then wife, she has been always there for me. If not for her, I would have been lost long before I completed this dissertation.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

In a traditional single-stage clinical trial, the data remain blinded until the conclusion of the study other than for data monitoring purposes. A clinical trial with an adaptive design, on the other hand, allows the opportunity to look at the data at an interim stage, and to modify the ongoing design based on observed and unblinded data. In this dissertation we introduce a framework for two-stage adaptive procedures, in which the design of the second stage depends on the first stage observations. We focus on unblinded examination of the first stage data, but in the literature both blinded and unblinded examination have been considered.

Blinded sample size re-estimation [e.g., Gould and Shih (1998)] does not break a blind at the end of Stage I, but uses techniques such as EM algorithm to estimate the pooled within-group variance. On the other hand, unblinded examination [e.g., Prochan and Hunsberger (1995)] reveals the group means and group variances at the end of the first stage and uses the information to design Stage II. And in between these two, there is "partial unblinding" technique in which only the variance estimate is revealed at the end of Stage I, but the group means are not revealed. [e.g., Zucker et. al. (1999)] In this dissertation blinded examination is not discussed, and the "partial unblinding" technique is briefly discussed in Chapter 3.

In the context of unblinded two-stage adaptive procedures in the literature, the phrase, "two-stage adaptive design" is used with two different notions of adaptiveness. There are two schools of thought as to how much adaptation is allowed at the end of Stage I. Some authors claim that one can design Stage II (i.e., choose sample size and critical value) after observing the Stage I data. These authors include Proschan and Hunsberger (1995) and Posch, Bauer and Brannath (2003). The others claim that it is necessary that the design of Stage II be defined for all possible Stage I outcomes prior to the start of Stage I [e.g., Liu and Chi (2001), Shun (2001)]. Then the design of Stage II is predetermined before Stage I, but it is not known which design (i.e., sample size, critical value and so on) will be realized because it depends on the Stage I observations. Predetermining the design of Stage II can permit the unconditional power of the procedure to be calculated prior to the start of

the study. We use the phrase "two-stage adaptive design" in this latter context. More discussion is given on the difference between the two ideas in Chapter 2.

As the title of this dissertation indicates, we develop a framework for two-stage adaptive procedures. In recent years, especially since the mid 1990's, there have been a number of papers published in the field of adaptive designs. Some propose new designs [e.g., Bauer and Köhne (1994), Lehmacher and Wassmer (1999)] while others compare existing procedures [e.g., Wassmer (1998), Posch and Bauer (1999), Hommel and Kropf(2001)]. It is not our goal to develop another procedure or to compare the existing procedures. Rather we develop a framework that facilitates the design of a two-stage adaptive procedure. Using our framework, it is possible to design a two-stage adaptive procedure which possesses certain design characteristics in terms of Type I and Type II error rates, sample sizes and so on. Additionally, the ability of our framework to place many, if not all, different two-stage adaptive procedures under a uniform formulation provides a tool to compare and contrast the previously proposed procedures in a systematic fashion.

In Chapter 2, we develop the framework in the most basic context, and show how the framework is used in designing a two-stage adaptive procedure using an example. A brief literature review is given of some previously proposed two-stage adaptive procedures by placing them under our formulation. We choose two specific procedures as examples and show in detail how these procedures fit into our framework. The first of the two examples is a procedure proposed by Lan and Trost (1995). We also show how our framework is used to facilitate a construction of similar two-stage design but with more favorable features in terms of the expected and maximum sample sizes. The second example is a group sequential procedure. We show how a two-stage group sequential procedure described in Jennison and Turnbull (1999) can be placed under our formulation. We also consider the monotonicity of the unconditional power function in a two-stage adaptive design. And we give the results of a computational experiment which compares different two-stage adaptive designs. This experiment is feasible because of the framework's ability to place different designs under one formulation.

In constructing the original framework in Chapter 2, our main focus is on sample size modification. The critical value of the Stage II rejection region and the conditional power of the Stage II test are also determined by Stage I data. However, from the viewpoint of the clinical trial sponsor, sample sizes are of interest because they directly influence the cost of the clinical trial. Therefore, whereas other modifications of goals, such as an adaptive choice of hypotheses [e.g., Hommel and Kropf (2001)] or changing the main efficacy endpoints [e.g., Bauer and Köhne (1994)] are considered in the literature, we focus on the sample size modification in the basic framework of Chapter 2.

In order to make the construction of the framework simple in Chapter 2, we make two key assumptions. The first one is that of balanced sample sizes from the treatment and control groups, and the second one is known variances. While these two assumptions keep algebraic complexity minimal, they may prevent us from applying this framework in practice. Therefore, in Chapter 3, we consider as extensions to the basic framework, situations in which these assumptions cannot be made. In Section 3.2, we show how to build a two-stage adaptive procedure with unbalanced sample sizes from two groups, and how to handle the design when some of the observations are missing completely at random in Stage I and/or in Stage II. In Section 3.3, we discuss a realistic situation in which the population variances are unknown at the beginning of the study and must be estimated within the study. Several authors, including Cui, Hung and Wang (1999) and Shun (2001), make the assumption of known variances in their respective papers. Unknown variances present little problem to authors [e.g., Lehmacher and Wassmer (1999), Kieser and Friede (2000) and Denne (2001)] who allow redesigning of Stage II after observing Stage I data, including a variance estimate. However, as stated before, our procedure predetermines the design of Stage II before Stage I for all possible outcomes of Stage I. Consequently, we need to establish a procedure that incorporates "updating the variance estimate" so that a design for Stage II is determined for each and all possible estimates of the population variances from Stage I.

In Chapter 4, we consider a practically motivated problem of switching the main study objectives in an active control study between one of testing noninferiority and one of testing superiority of the treatment. An active control study, unlike a placebo control study, is a clinical trial in which a new treatment is compared to a traditional treatment whose efficacy has been well established. Such studies have become increasingly popular in recent years, especially in research areas related to oncology and HIV. A traditional placebo control study is often impossible to conduct in these and similar areas, because it may not be ethical to require a group of patients to receive no effective treatment. Superiority trials and noninferiority trials are two types of active control studies. The former is a trial to show that the new treatment is better than the active control, and the latter seeks to show that the new treatment is at least as good as the active control.

A clinical trial with an adaptive design may start with noninferiority as its main objective. At the end of Stage I, however, the data may show strong evidence that concluding superiority may be plausible. Then it may be desirable to switch the main objective to superiority and adjust the Stage II sample size for the new goal of the clinical trial. On the other hand, a clinical trial may start with the main objective to show superiority, and at the end of Stage I, the data may suggest that showing superiority is very difficult and costly. In such a situation, an adaptive design may switch the main objective to noninferiority, and in doing so reduce the necessary sample size.

A number of papers [e.g., CPMP (2000)] consider testing noninferiority and superiority simultaneously in a single-stage clinical trial; these papers are discussed in the literature review section of Chapter 4. There are also a few papers [e.g., Wang, et.al. (2001) and Brannath, et.al. (2003)] which combine two-stage adaptive designs and notions concerning active control studies to permit the hypotheses of interests to be switched at the interim stage. These papers are discussed also in the literature review section of Chapter 4.

Two-stage adaptive procedures are relatively new as a field of research. Although this dissertation gives a tool to design and analyze a two-stage adaptive procedure, there are still questions yet to be answered. In Chapter 5, we summarize issues and topics for future research in the field of two-stage adaptive procedures.

Chapter 2

The Basic Framework For Two-Stage Adaptive Procedure

2.1 Introduction

We introduce in this chapter a framework for two-stage adaptive procedure. Unlike a traditional single-stage clinical trial in which the data remain blinded until the conclusion of the study, a clinical trial with a two-stage adaptive design breaks the blind at the end of Stage I and allows modification of the ongoing design in Stage II in light of observed data.

In the current chapter, we only consider the most fundamental situation in which only the basic features of Stage II, such as the sample size and the critical value, depend on Stage I observations. More practically interesting problems such as updating variance estimates, incorporating missing observations and switching the main study objectives, will be presented in later chapters as extensions of the basic framework that is developed in this chapter.

The motivation for developing a two-stage adaptive procedure stems from the fact that it is difficult to prespecify some of the design parameters prior to the beginning of the study. When the design parameters are incorrectly prespecified in a traditional single stage study, the sample size will be either too small or too large for the clinical trial to have the claimed power. It is considered unethical for a clinical statistician to recommend designs using more patients than necessary. It is also unethical to design an underpowered clinical trial in which the probability of reaching a positive conclusion is low, and thus patients' participation may put them at risk with low likelihood of a positive societal outcome.

A clinical trial with an adaptive design gives investigators the opportunity to look at the data at an interim stage. The design of Stage II depends on the observations from Stage I. It will be shown in the current chapter that a two-stage adaptive design may present an advantage in terms of expected sample size over the conventional single-stage design. The challenge for statistical methodology is to maintain proper control of experimental error rates when permitting adaptation.

Proschan and Hunsberger (1996) introduced the concept of a "Conditional Error Function." With their formulation, we can formulate an adaptive design which gives the clinical trial sponsor a chance to increase the sample size and protects the Type I error rate. There have been a number of papers published in this field. Some introduce new two-stage adaptive procedures [e.g., Bauer and Köhne (1994), Lehmacher and Wassmer (1999)], while others compare existing procedures [e.g., Posch and Bauer (1999) and Hommel and Kropf (2001)]. As noted by many authors, different two-stage adaptive procedures can be placed under the same formulation using different conditional error functions. There are also a number of papers addressing issues related to two-stage adaptive designs such as inflation of the Type I error rate, the conditional and unconditional power and operational difficulties when implementing an adaptive design. Among such papers are those by Gould (2001), Denn (2001) and Shun (2001). What is lacking in this field is a unified framework for the adaptive designs. As will be shown by examples, different procedures can be placed under the framework introduced in this chapter. With our framework, we can study the properties of two-stage adaptive procedures in an organized fashion.

The most important contribution of our framework is its ability to facilitate the design of two-stage adaptive procedures. We show that in a two-stage adaptive procedure, there are nine components that specify a design uniquely. We provide a calculus to manipulate these nine components so that the resulting design possesses certain characteristics.

In Section 2.2, the framework is established. Our framework is defined by nine specification components (five for Stage I and four for Stage II.) For Stage I, these components are:

- α_1, the probability of rejecting H_0 when the null hypothesis is true.
- β_1, the probability of accepting H_0 under a specified alternative hypothesis.
- n_1, the Stage I sample size.
- k_1, the Stage I critical value.
- k_2, the Stage I critical value.

For Stage II, there are 4 specification components, namely:

- $A(y_1, \xi_0)$, the conditional probability of rejecting the null hypothesis when the null hypothesis is true,
- $A(y_1, \xi_1)$, the conditional probability of rejecting the null hypothesis under the specified alternative hypothesis,

- $n_2(y_1)$, the Stage II sample size,
- $w(y_1)$, the Stage II critical value,

all of which are expressed as functions of the summary observation y_1 from Stage I. We show how previously proposed two-stage adaptive procedures (including the group sequential method) fit into our framework.

In Section 2.3, we consider monotonicity of the power function in a two-stage adaptive procedure. This topic has been apparently overlooked in the field of two-stage adaptive procedures. Without the property of monotonicity of the power function, the level of significance of the test cannot be insured by merely controlling its probability of the Type I error at the boundary between the null and alternative hypotheses. We give simple conditions that guarantee that the power function is increasing.

In Section 2.4, we demonstrate, using two examples, the flexibility of our framework in constructing a two-stage adaptive procedure. In Section 2.4.1, we show how a two-stage group sequential method can be placed under our framework. And in Section 2.4.2, we consider a procedure proposed by Lan and Trost (1996). First we show that their procedure, too, can be placed under our framework. Then we use our framework's ability to facilitate a design to effect a substantial improvement to Lan and Trost's design.

A question that has been asked a number of times concerns how to optimally choose a two-stage adaptive procedure. So far, a clear answer to this question has not been given. In Section 2.5, we use an ANOVA analysis to compare designs indexed by certain design components to see which of these components are most influential in determining the performance characteristics of the designs. Having a unified framework enables us to compare different designs in this fashion. In order to compare the performance characteristics of different designs, we use the following three performance criteria. The first criterion is the power at points intermediate between the null hypothesis and the original alternative. All designs that we consider have a fixed Type I error rate and fixed power at the original alternative. Thus, all the differences in the power curve occur at the intermediate points between the null and the alternative values. The second and the third performance criteria are both functions of the expected sample sizes. These criteria are introduced and explained in detail in Section 2.5. Although the optimal design is not given, we provide the following conclusion based on the ANOVA analysis. In determining characteristics of the design in terms of the power and expected sample size, the Stage I design components, n_1, α_1 and β_1 are far more important than such Stage II design components as the shape of the conditional power functions.

2.2 Framework

2.2.1 Construction

We suppose independent sampling from two normal populations with respective means μ_t and μ_c and respective variances σ_t^2 and σ_c^2. So we have

$$X_t \sim \text{Normal}\left(\mu_t,\, \sigma_t^2\right),$$
$$X_c \sim \text{Normal}\left(\mu_c,\, \sigma_c^2\right).$$

Consider testing

$$H_0 : \mu_t - \mu_c \leq \Delta_0$$
$$H_1 : \mu_t - \mu_c > \Delta_0$$

where the subscripts stand for "treatment" and "control", respectively. Both superiority and non-inferiority hypotheses can be written in this form.

In some of the recent research on two-stage adaptive designs, the null hypothesis is simple, i.e., $H_0 : \mu_t - \mu_c = \Delta_0$ [e.g., Proschan and Hunsberger (1995), Lan and Trost (1997), Jennison and Turnbull (1999)]; however, we consider the more appropriate composite null hypothesis. When the probability of the Type I error is only protected for a simple $H_0 : \mu_t - \mu_c = \Delta_0$, the probability of rejecting H_0 when $\mu_t - \mu_c < \Delta_0$ can be higher than the specified α. What enables us to control the Type I error probabilities for the composite null hypothesis in the adaptive designs that we consider is the monotonicity of the power function. In this section, we simply assume that the power function is monotone without proof; later in Section 2.3, we show that the power functions of the two-stage adaptive procedures that we consider are monotone.

Consider a two-stage adaptive design where in Stage I, n_{1t} observations are taken from the treatment group and n_{1c} from the control group. In order to make the construction of the framework simple, we make the assumption that $n_{1t} = n_{1c} \equiv n_1$. Later in Section 3.2, we consider the situations in which the sample sizes from the treatment and control groups are different. The distributions of the sample means for Stage I are:

$$\bar{X}_{1t} \sim \text{Normal}\left(\mu_t,\, \frac{\sigma_t^2}{n_1}\right),$$
$$\bar{X}_{1c} \sim \text{Normal}\left(\mu_c,\, \frac{\sigma_c^2}{n_1}\right).$$

Let $\sigma^2 = \dfrac{\sigma_t^2 + \sigma_c^2}{2}$ and define

$$Y_1 = \frac{\bar{X}_{1t} - \bar{X}_{1c}}{\sqrt{2}\,\sigma}. \tag{2.2.1}$$

For the sake of simplicity, we assume that σ_t^2 and σ_c^2 are known. The situation in which the population variances are unknown is treated in Section 3.3.

Note that Y_1 is a test statistic based upon unblinded data. We have the following result:

$$Y_1 \sim \text{Normal}\left(\frac{\Delta}{\sqrt{2}\,\sigma},\, \frac{2\sigma^2}{2\sigma^2 n_1}\right) = \text{Normal}\left(\xi,\, \frac{1}{n_1}\right), \tag{2.2.2}$$

where $\xi \equiv \frac{\Delta}{\sqrt{2}\,\sigma} = \frac{\mu_t - \mu_c}{\sqrt{2}\,\sigma}$. Although we set up the problem for comparison of two treatments, the above simple formulation allows us also to apply this method to comparison of one treatment to a constant.

A decision is made at the end of Stage I according to the following decision rule.

$$\text{if } \begin{cases} Y_1 \in (-\infty,\, k_1] & \text{stop and accept } H_0, \\ Y_1 \in (k_1,\, k_2] & \text{continue to Stage II}, \\ Y_1 \in (k_2,\, \infty) & \text{stop and reject } H_0. \end{cases}$$

In Stage II, additional samples of size $n_2(y_1)$ are taken from each of the control and treatment groups. As the notation indicates, the Stage II sample size is permitted to be a function of y_1. Let $\bar{X}_{2t}$ and $\bar{X}_{2c}$ be the means of the observations in Stage II from the treatment and control groups, respectively, and define

$$Y_2 = \frac{\bar{X}_{2t} - \bar{X}_{2c}}{\sqrt{2}\,\sigma}. \tag{2.2.3}$$

The Stage I observations influence the distribution of Y_2 only through the dependence of the sample size $n_2(y_1)$ on y_1, so that given that $Y_1 = y_1$,

$$Y_2 \sim \text{Normal}\left(\xi,\, \frac{1}{n_2(y_1)}\right).$$

At the end of Stage II,

$$\text{if } \begin{cases} Y_2 \in (-\infty,\, w(y_1)] & \text{stop and accept } H_0, \\ Y_2 \in (w(y_1),\, \infty) & \text{stop and reject } H_0. \end{cases}$$

The Stage II critical value, $w(y_1)$, is a function of y_1.

The decision rule at the end of Stage II is based on Y_2, but clearly we are not disregarding the information from the Stage I observations. All the information from Stage I is embedded into $n_2(y_1)$ and $w(y_1)$. Equivalently, we can express the Stage II decision rule in terms of Y_b, the test statistic for the entire study,

$$Y_b = \frac{n_1 Y_1 + n_2(y_1) Y_2}{n_b(y_1)}, \tag{2.2.4}$$

where $n_b(y_1) = n_1 + n_2(y_1)$. Then the conditional distribution of Y_b given $Y_1 = y_1$ is

$$Y_b \sim \text{Normal}\left(\frac{n_1 y_1 + n_2(y_1)\xi}{n_b(y_1)}, \frac{n_2(y_1)}{(n_b(y_1))^2}\right).$$

H_0 is rejected at the end of Stage II if $Y_b > c(y_1)$, and is accepted otherwise. From the relation between Y_2 and Y_b, we have the following connections:

$$\begin{aligned} w(y_1) =& \frac{n_b(y_1)c(y_1) - n_1 y_1}{n_2(y_1)}, \\ c(y_1) =& \frac{n_1 y_1 + n_2(y_1)w(y_1)}{n_b(y_1)}. \end{aligned} \tag{2.2.5}$$

Let α be the prespecified overall Type I error rate at Δ_0 (equivalently at $\xi_0 = \Delta_0/\sqrt{2}\,\sigma$), and let β be the prespecified overall Type II error rate at the prespecified trial design alternative Δ_1 (equivalently at $\xi_1 = \Delta_0/\sqrt{2}\,\sigma$) with corresponding overall power ρ, i.e., $\rho = 1 - \beta$. We assume that $\alpha < .5$ and $\beta < .5$, and define the Type I error rate, Type II error rate and power for Stage I as follows:

$$\begin{aligned} \alpha_1 =& P[\,\text{Reject } H_0 \text{ in Stage I} \,|\, \xi = \xi_0], \\ \beta_1 =& P[\,\text{Accept } H_0 \text{ in Stage I} \,|\, \xi = \xi_1], \\ \rho_1 =& P[\,\text{Reject } H_0 \text{ in Stage I} \,|\, \xi = \xi_1]. \end{aligned}$$

We assume that α and β are prespecified for any two-stage adaptive procedure we consider.

For Stage II, define the "conditional power function," $A(y_1, \xi^*)$, for $\xi^* \in (-\infty, \infty)$ as follows:

$$A(y_1, \xi^*) = P_{\xi^*}[\,\text{Reject } H_0 \text{ in Stage II } \,|\, Y_1 = y_1\,], \tag{2.2.6}$$

where $P_{\xi^*}[\ ]$ is the probability under $\xi = \xi^*$. This definition of $A(y_1, \xi^*)$ is an extension of the "conditional error function," introduced by Proschan and Hunsberger (1995). Their conditional error function is the conditional power function evaluated at $\xi^* = \xi_0$. When $y_1 \le k_1$, H_0 is accepted in Stage I, and when $k_2 < y_1$, H_0 is rejected in Stage I. It thus follows that $A(y_1, \xi^*) = 0$ for $y_1 \le k_1$ and $A(y_1, \xi^*) = 1$ for $k_2 < y_1$. The conditional Type I error rate given y_1 is $A(y_1, \xi_0)$, and the conditional power at the original alternative, ξ_1, given y_1 is $A(y_1, \xi_1)$. Because $A(y_1, \xi^*)$ is the *conditional* power function after observing y_1, ξ^* can be allowed to depend on y_1. For example, $A(y_1, \xi^*(y_1)) = A(y_1, y_1)$ is the conditional power function when the alternative is updated after Stage I to be the Stage I effect size, y_1.

To describe the Stage I design components of our procedure, we need to provide the following five components.

(a) α_1 $\cdots$ the probability of rejecting H_0 when ξ_0 is true,

(b) β_1 $\cdots$ the probability of accepting H_0 when the alternative, ξ_1, is true,

(c) n_1 $\cdots$ the Stage I sample size,

(d) k_1 $\cdots$ the Stage I critical value,

(e) k_2 $\cdots$ the Stage I critical value.

For Stage II, there are four specification components, namely:

(i) $A(y_1, \xi_0)$ $\cdots$ the conditional power when ξ_0 is true,

(ii) $A(y_1, \xi_1)$ $\cdots$ the conditional power function when the alternative, ξ_1, is true,

(iii) $n_2(y_1)$ $\cdots$ the Stage II sample size function,

(iv) $w(y_1)$ $\cdots$ the Stage II critical value function.

There are some conditions that these design components must satisfy. Obviously, n_1 and $n_2(y_1)$ must be positive integers, and $A(\bullet, \bullet)$ must be a non-negative function which range in $[0, 1]$. We also need $\alpha_1 < \alpha$, and $\beta_1 < \beta$. Furthermore, in order to have the overall Type I error rate equal to α and the overall power under ξ_1 equal to ρ, $A(y_1, \xi_0)$ and $A(y_1, \xi_1)$ must satisfy the following conditions.

$$\alpha_2 \equiv \alpha - \alpha_1 = \int_{k_1}^{k_2} A(y_1, \xi_0) g_{\xi_0}(y_1)\, dy_1, \tag{2.2.7}$$

$$\rho_2 \equiv \rho - \rho_1 = \int_{k_1}^{k_2} A(y_1, \xi_1) g_{\xi_1}(y_1)\, dy_1, \tag{2.2.8}$$

where $g_\xi(y_1)$ is the pdf of Normal $(\xi,\, 1/n_1)$. In addition $A(y_1, \xi_0)$ and $A(y_1, \xi_1)$ must satisfy the technical condition of being measurable in y_1 as argued by Liu, Proschan and Pledger (2002).

Any two-stage adaptive design can be described in terms of these nine specification components: (a) - (e) for Stage I and (i) - (iv) for Stage II. These nine components have functional relationships with one another and cannot be determined independently of the others. Because the underlying distribution of Y_1 is (2.2.2), we have the following relation between α_1, n_1 and k_2.

$$\begin{aligned} \alpha_1 =& P_{\xi_0}\left[Y_1 > k_2\right] \\ =& 1 - \Phi\left[\sqrt{n_1}\,(k_2 - \xi_0)\right]. \end{aligned}$$

Therefore,

$$z_{\alpha_1} = \sqrt{n_1}\,(k_2 - \xi_0), \tag{2.2.9}$$

where $z_a = \Phi^{-1}(1-a)$. Similarly, we have the following relation between β_1, n_1 and k_1.

$$\begin{aligned}\beta_1 =& P_{\xi_1}\left[Y_1 < k_1\right] \\ =& \Phi\left[\sqrt{n_1}\ (k_1 - \xi_1)\right].\end{aligned}$$

Therefore,

$$z_{\beta_1} = -\sqrt{n_1}\ (k_1 - \xi_1). \tag{2.2.10}$$

Thus given two of α_1, n_1 and k_2, the third quantity is determined from (2.2.9), and given two of β_1, n_1 and k_1, the third quantity can be found from (2.2.10). Additional care must be taken in singular cases, e.g., $\alpha_1 = 0$. Note that these two groupings of design components are connected through the common component, n_1. We summarize the connection between these five components as follows:

Fact 2.2.1. *Any three specifications of the components of (a) - (e) chosen so that there is at least one component from each of {(a), (c), (e)} and {(b), (c), (d)} are sufficient to determine the other two.*

The three components that determine a Stage I design must be chosen with care. The resulting procedure must satisfy the aforesaid conditions that $\alpha_1 < \alpha$, $\beta_1 < \beta$ and n_1 is a positive integer. While it is possible to specify more than three components, our choices must be coherent so that (2.2.9) and (2.2.10) have a solution.

Given $Y_1 = y_1$ such that $k_1 < y_1 \leq k_2$, we have an analogous situation for Stage II, but with only one critical value function, $w(y_1)$, instead of the two critical values, k_1, k_2 in Stage I.

Fact 2.2.2. *Any two specifications of the components of (i) - (iv) are sufficient to determine the other two.*

The connections for these four Stage II design components are summarized as follows:

$$\begin{aligned}A(y_1, \xi_0) =& P_{\xi_0}\left[\,\text{Reject } H_0 \text{ in Stage II} \,|\, Y_1 = y_1\,\right] \\ =& P_{\xi_0}\left[\,Y_2 > w(y_1) \,|\, Y_1 = y_1\,\right] \\ =& 1 - \Phi\left[\sqrt{n_2(y_1)}\ (w(y_1) - \xi_0)\right].\end{aligned} \tag{2.2.11}$$

Therefore,

$$z_{A(y_1,\xi_0)} = \sqrt{n_2(y_1)}\,(w(y_1) - \xi_0). \tag{2.2.12}$$

Similarly,

$$\begin{aligned} A(y_1, \xi_1) =& P_{\xi_1} \left[\text{Reject } H_0 \text{ in Stage II} \,|\, Y_1 = y_1 \right] \\ =& 1 - \Phi \left[\sqrt{n_2(y_1)} \, (w(y_1) - \xi_1) \right]. \end{aligned}$$

Therefore,

$$z_{A(y_1,\xi_1)} = \sqrt{n_2(y_1)} \, (w(y_1) - \xi_1). \tag{2.2.13}$$

Using (2.2.12) and (2.2.13), we have that any two specification components from (i) $A(y_1, \xi_0)$, (ii) $A(y_1, \xi_1)$, (iii) $n_2(y_1)$ and (iv) $w(y_1)$ are sufficient to solve for the other two. If $A(y_1, \xi_0)$ and $A(y_1, \xi_1)$ are given, we can solve for $n_2(y_1)$ and $w(y_1)$ as follows.

$$n_2(y_1) = \frac{(z_{A(y_1,\xi_0)} - z_{A(y_1,\xi_1)})^2}{(\xi_1 - \xi_0)^2}, \tag{2.2.14}$$

$$w(y_1) = \frac{z_{A(y_1,\xi_0)}\xi_1 - z_{A(y_1,\xi_1)}\xi_0}{z_{A(y_1,\xi_0)} - z_{A(y_1,\xi_1)}}. \tag{2.2.15}$$

On the other hand, if $n_2(y_1)$ and $w(y_1)$ are given, we can solve for $A(y_1, \xi)$ for any ξ by

$$A(y_1, \xi) = 1 - \Phi \left[\sqrt{n_2(y_1)} \, (w(y_1) - \xi) \right]. \tag{2.2.16}$$

Or if $A(y_1, \xi^*)$ and $n_2(y_1)$ are given, we can solve for $A(y_1, \xi)$ for any ξ and $w(y_1)$ by

$$\begin{aligned} A(y_1, \xi) =& 1 - \Phi \left[z_{A(y_1,\xi^*)} + \sqrt{n_2(y_1)} \, (\xi^* - \xi) \right], \\ w(y_1) =& \frac{z_{A(y_1,\xi^*)}}{\sqrt{n_2(y_1)}} + \xi^*. \end{aligned} \tag{2.2.17}$$

Or if $A(y_1, \xi^*)$ and $w(y_1)$ are given, we can solve for $A(y_1, \xi)$ for any ξ and $n_2(y_1)$ by

$$\begin{aligned} A(y_1, \xi) =& 1 - \Phi \left[\frac{w(y_1) - \xi}{w(y_1) - \xi^*} \, z_{A(y_1,\xi^*)} \right], \\ w(y_1) =& \left(\frac{z_{A(y_1,\xi^*)}}{w(y_1) - \xi^*} \right)^2. \end{aligned} \tag{2.2.18}$$

Whatever two components we choose must be chosen so that the resulting procedure satisfies the conditions that $n_2(y_1)$ is a positive integer for all $y_1 \in (k_1, k_2]$, and also that the conditions on $A(y_1, \xi_0)$ and $A(y_1, \xi_1)$ stated in (2.2.7) and (2.2.8) hold. Although $n_2(y_1)$ must be integer-valued, in order to simplify computations we relax the condition and allow $n_2(y_1)$ to be any positive, real number for all $y_1 \in (k_1, k_2]$. The actual sample size, then, will be the smallest integer that is equal to or greater than $n_2(y_1)$.

Our choice of Stage II design components can be made more flexible. Let $\xi^*(y_1)$ and $\xi^\dagger(y_1)$ be suitable functions of y_1 (with additional technical conditions noted below) and consider (i*)

$A(y_1, \xi^*(y_1))$ and (ii*) $A(y_1, \xi^\dagger(y_1))$. Then the components (i) and (ii) of Fact 2.2.2 can be replaced by the preceding (i*) and (ii*). Even when we use the preceding specification of components (i*) and (ii*), conditions (2.2.7) and (2.2.8) must be satisfied; to check these we need to determine $A(y_1, \xi_0)$ and $A(y_1, \xi_1)$ from (i*) and (ii*). We can do so from the following result.

Fact 2.2.3. *Given $A(y_1, \xi^*(y_1))$ and $A(y_1, \xi^\dagger(y_1))$, any $A(y_1, \xi)$ can be derived by*

$$A(y_1, \xi) = 1 - \Phi\left[\frac{\xi^\dagger(y_1) - \xi}{\xi^\dagger(y_1) - \xi^*(y_1)} z_{A(y_1,\xi^*(y_1))} + \frac{\xi - \xi^*(y_1)}{\xi^\dagger(y_1) - \xi^*(y_1)} z_{A(y_1,\xi^\dagger(y_1))}\right]. \tag{2.2.19}$$

Some conditions should be noted. First, clearly (2.2.19), as well as (2.2.14) and (2.2.15), with ξ_0 and ξ_1 replaced by $\xi^*(y_1)$ and $\xi^\dagger(y_1)$, do not have solutions when $\xi^*(y_1) = \xi^\dagger(y_1)$. For technical reasons, we need to require that $\xi^*(y_1) = \xi^\dagger(y_1)$ only for a finite number of points. Pragmatically, because there are a finite number of points, the probability of observing any of these points is zero. For esthetics, one could obtain the values of $A(\bullet, \xi)$, $n_2(\bullet)$ and $w(\bullet)$ at these points by imposing continuity conditions on $A(y_1, \xi)$, $n_2(y_1)$ and $w(y_1)$.

Another caution is necessary when $A(y_1, \xi^*)$ and $w(y_1)$ are specified. From the relation between these two components, $A(y_1, \xi^*) = 0.5 \Leftrightarrow w(y_1) = \xi^*$. At any y_1 where $A(y_1, \xi^*) = 0.5$ and $w(y_1) = \xi^*$, the other two components cannot be determined. We impose continuity conditions again to solve for $n_2(y_1)$ and $A(y_1, \xi)$ at $\xi \neq \xi^*$. Therefore, in this special case, we need to require that $A(y_1, \xi^*) = 0.5$ (equivalently, $w(y_1) = \xi^*$) only for a finite number of y_1.

Finally, when two specifying components are $A(y_1, \xi^*)$ and $A(y_1, \xi^\dagger)$, we need to make sure that $A(y_1, \xi^*) > A(y_1, \xi^\dagger)$ if and only if $\xi^* > \xi^\dagger$ because $A(y_1, \xi)$ is an increasing function in ξ given y_1.

2.2.2 Previously Proposed Procedures

A number of two-stage adaptive procedures have been proposed in the literature. They can be expressed in terms of the framework of Section 2.2.1 with different specification components for Stage I and for Stage II. The following summary gives the specifications of the components for various proposed procedures. Posch and Bauer (1999) and Hommel and Kropf (2001) give summaries of some of these procedures by placing them into one formulation using the conditional error functions. Wassmer (1999) makes comparison of the procedures of Proschan and Hunsberger (1995) and Bauer and Köhne (1994) and finds that the behaviors of these procedures in terms of power are very similar.

Some of these noted procedures use the Stage II p-value to make a decision at the end of Stage II. The following lemma connects the decision rule defined by the Stage II p-value and the decision rule defined by the Stage II critical value, $w(y_1)$.

Lemma 2.2.1. *In a two stage adaptive procedure expressed in terms of the design components of our framework, the following two decision rules at the end of Stage II are equivalent. (1) Reject H_0 if $y_2 > w(y_1)$, and (2) reject H_0 if Stage II p-value $< A(y_1, \xi_0)$ evaluated at the observed y_1.*

Proof: Suppose that we observe $Y_2 = y_2$ in Stage II. Because under the null,

$$Y_2 \sim \text{Normal}\left(\xi_0, \frac{1}{\sqrt{n_2(y_1)}}\right),$$

the Stage II p-value can be written as

$$p_2 \equiv P\left[Y_2 > y_2\right] = 1 - \Phi\left[\sqrt{n_2(y_1)}\,(y_2 - \xi_0)\right].$$

And from (2.2.11),

$$A(y_1, \xi_0) = 1 - \Phi\left[\sqrt{n_2(y_1)}\,(w(y_1) - \xi_0)\right].$$

Thus

$$\begin{aligned} p_2 < A(y_1, \xi_0) &\Leftrightarrow 1 - \Phi\left[\sqrt{n_2(y_1)}\,(y_2 - \xi_0)\right] < 1 - \Phi\left[\sqrt{n_2(y_1)}\,(w(y_1) - \xi_0)\right] \\ &\Leftrightarrow \sqrt{n_2(y_1)}\,(y_2 - \xi_0) > \sqrt{n_2(y_1)}\,(w(y_1) - \xi_0) \\ &\Leftrightarrow y_2 > w(y_1). \end{aligned}$$

□

We give in Table 2.1, the specification components of the various previously proposed two-stage adaptive procedures, where the symbol "•" indicates a specification component and the symbol "-" indicates a specification component which is implied but not fully defined.

Table 2.1: *Specification components of previously proposed adaptive procedures.*

	n_1	α_1	β_1	k_1	k_2	$A(y_1, \xi_0)$	$A(y_1, \xi^*)$	$n_2(y_1)$	$w(y_1)$
(a) PH	•		-	•	-	•			
(b) BK	-	•		•		•			
(c) GSP	•			•	•			•	•
(d) LW	•			•	•			-	•
(e) LC	•	•		•		•	•		
(f) LD	-	•	•			•		•	

PH = Proschan and Hunsberger (1996). BK = Bauer and Köhne (1994).
GSP = group sequential procedure. LW = Lehmacher and Wassmer (1999).
LC = Liu and Chi (2001). LD = Lan and DeMets (1983).

We now provide detailed discussion of the procedures described in Table 2.1.

(a) Proschan and Hunsberger (1996) in defining their procedure use $n_1 = (z_\alpha + z_\beta)^2/(\xi_1 - \xi_0)^2$, the original sample size for the conventional one-stage design with the same α and β. They specify a particular $A(y_1, \xi_0)$ for Stage II, which is $A(y_1, \xi_0) = 1 - \Phi\left[\sqrt{n_1}\sqrt{(k_2 - \xi_0)^2 - (y_1 - \xi_0)^2}\right]$. This $A(y_1, \xi_0)$ is not flexible; that is, we cannot choose any set of k_1 and k_2 to use with this $A(y_1, \xi_0)$. In order to protect the Type I error rate, (2.2.7) must be satisfied. Consequently, for Proschan and Hunsberger's $A(y_1, \xi_0)$, fixing k_2 automatically fixes k_1 through (2.2.7). For their procedures, Proschan and Hunsberger give functional relationships between the conditional power, Stage II sample size and Stage II critical value. They state that for an observed y_1, conditional power can be chosen subjectively, and that this choice will determine the Stage II sample size and critical value. As is discussed at the end of this section, it is necessary that these elements be defined before observing y_1 for all values of $y_1 \in [k_1, k_2)$. If Proschan and Hunsberger had provided the functions $A(y_1, \xi_1)$, $n_2(y_1)$ and $w(y_1)$ beforehand, their procedure would have been in our framework.

(b) Bauer and Köhne (1994) specify α_1 and P(Accept H_0 in Stage I $\mid H_0$), which determines k_1 if n_1 is given. For Stage II, the rejection region is given in terms of the conditional p-value, which by Lemma 2.2.1, is equivalent to specifying $A(y_1, \xi_0)$. The procedure of Bauer and Köhne lacks one of two specification components for Stage II. The discussion at the end of this section, therefore, also applies to their procedure.

(c) Our framework encompasses the conventional one-sided two-stage group sequential tests. [Jennison and Turnbull (1999) discuss these tests.] For two-stage group sequential tests, the sample sizes for both stages (assumed equal) and the critical values for both stages are given in advance. Therefore, specification components for this are n_1, k_1 and k_2 for Stage I; and $n_2(y_1) = n_1$ and $w(y_1)$ obtained through $c(y_1) = c$ [see (2.2.5)]. The connection between the group sequential method and our framework is discussed in detail in Section 2.4.1.

(d) Lehmacher and Wassmer's (1999) procedure, like most group sequential methods, is essentially designed for more than two stages. When artificially restricted to two stages, this procedure can be placed in our framework. For Stage I, specification components are n_1, k_1 and k_2. The critical values are the same as for a two-stage group sequential test. If $n_2(y_1) = n_1$ for all $y_1 \in [k_1, k_2)$, this procedure would be identical to the group sequential method described in (c) above. However, without an explicit rule to determine the Stage II sample size, it lacks one specification component for Stage II.

(e) Liu and Chi (2001) specify n_1, α_1 and β_1 for Stage I. For Stage II, they specify $A(y_1, \xi_0) = 1 - \Phi\left[\gamma\sqrt{n_1}\sqrt{(k_2 - \xi_0)^2 - (y_1 - \xi_0)^2}\right]$, which is an extension of the conditional power function that Proschan and Hunsberger use, and $A(y_1, \xi^*) = \rho_2$, where ξ^* is the minimum significant effect size, and ρ_2 is a conditional probability of rejecting H_0 in Stage II, which does not depend on y_1.

(f) Lan and DeMets' (1983) concept of alpha-spending is suitable for more than two stages, but two-stage versions fit into our framework. In alpha-spending, values α_1 and $A(y_1, \xi_0) = \alpha - \alpha_1$ for all $y_1 \in [k_1, k_2)$ are specified. Because their method does not allow early termination in favor of H_0, it follows that $\beta_1 = 0$. In their formulation, the Stage II sample size is equal to n_1 regardless of the observed Stage I value, y_1. Thus the specification components are α_1, β_1 for Stage I and $n_2(y_1) = n_1$, $A(y_1, \xi_0) = \alpha - \alpha_1$ for Stage II.

Lan and Trost (1997) also propose a two-stage adaptive procedure. A detailed discussion of this design is given in Section 2.4.2.

As summarized above, some of the procedures proposed in the literature do not explicitly define the components for Stage II as a function of y_1. Allowing Stage II components to be determined after y_1 is observed appears to offer flexibility. However Liu, Proschan and Pledger (2002) argue that for any two stage adaptive design, $n_2(y_1)$ must be a measurable function of y_1, as must be $w(y_1)$, $A(y_1, \xi_0)$ and $A(y_1, \xi_1)$. They demonstrate that one is not permitted to choose the conditional power and/or the Stage II sample size after observing a particular y_1, rather than prespecifying conditional power and Stage II sample size for all $y_1 \in (k_1, k_2]$. Beyond the specific measurability requirement, we need the specification of the noted functions of y_1 in the design phase for all $y_1 \in (k_1, k_2]$ in order to be able to calculate, when planning the design, the overall unconditional power at a specified alternative. Therefore, it is important that these components for Stage II be defined before observing y_1 in Stage I.

2.3 Power Function Considerations

We define the conditional power function in (2.2.6). Now we consider the *unconditional* power function. We allow for the calculation of conditional power at an alternative, ξ^*, depending on y_1. However, in order to define the unconditional power, we need to specify a ξ^* that is fixed and does not depend on y_1.

Suppose that a two-stage adaptive design is defined using Fact 2.2.1 and Fact 2.2.2; that is, whichever specification components are used, all nine (five for Stage I and four for Stage II) specification components are properly specified. Then we can write the unconditional power function, $\psi(\xi)$ for $\xi \in (-\infty, \infty)$ as follows:

$$\psi(\xi) = \int_{-\infty}^{\infty} A(y_1, \xi) g_\xi(y_1)\, dy_1, \tag{2.3.20}$$

where $g_\xi(y_1)$ is the pdf of Normal $(\xi,\, 1/n_1)$.

Because we consider a composite null hypothesis of the form, $H_0 : \Delta \leq \Delta_0$, the power function, $\psi(\xi)$ needs to be bounded above by α for all values of Δ below ξ_0. A straightforward way to

guarantee that the power function is bounded above for all ξ in $(-\infty, \xi_0]$ is the monotonicity of the power function in this range of ξ. The following lemma gives a sufficient condition for the power function (2.3.20) to be nondecreasing over an interval (a, b).

Lemma 2.3.1. *A sufficient condition for the unconditional power function, $\psi(\xi)$, to be nondecreasing in $\xi \in (a, b)$ is that $A(y_1, \xi)$ is nondecreasing in y_1 for every fixed $\xi \in (a, b)$.*

Proof: We need to show that $\psi(\xi_3) \geq \psi(\xi_2)$ for arbitrary ξ_2 and ξ_3 such that $\xi_3 > \xi_2$. We have

$$\begin{aligned}
\psi(\xi_3) &= \int_{-\infty}^{\infty} A(y_1, \xi_3) g_{\xi_3}(y_1) dy_1 \\
&= \int_{-\infty}^{\infty} A(y_1, \xi_3) \frac{g_{\xi_3}(y_1)}{g_{\xi_2}(y_1)} g_{\xi_2}(y_1) dy_1 \\
&= \int_{-\infty}^{\infty} A(y_1, \xi_3) g_{\xi_2}(y_1) dy_1 + \mathrm{Cov}\left[A(Y_1, \xi_3), \frac{g_{\xi_3}(Y_1)}{g_{\xi_2}(Y_1)} \right].
\end{aligned}$$

For a fixed $y_1 \in (k_1, k_2]$, by the definition of $A(y_1, \xi)$ in (2.2.6),

$$A(y_1, \xi_3) \geq A(y_1, \xi_2) \text{ for all } y_1 \in (-\infty, \infty).$$

Moreover, by the given condition, $A(y_1, \xi_3)$ is nondecreasing in y_1, and because normal densities have monotone likelihood ratio, $\dfrac{g_{\xi_3}(y_1)}{g_{\xi_2}(y_1)}$ is nondecreasing in y_1. Then it follows that

$$\mathrm{Cov}\left[A(Y_1, \xi_3), \frac{g_{\xi_3}(Y_1)}{g_{\xi_2}(Y_1)} \right] \geq 0.$$

Therefore,

$$\begin{aligned}
\psi(\xi_3) &\geq \int_{-\infty}^{\infty} A(y_1, \xi_2) g_{\xi_2}(y_1) dy_1 + \mathrm{Cov}\left[A(Y_1, \xi_3), \frac{g_{\xi_3}(Y_1)}{g_{\xi_2}(Y_1)} \right] \\
&\geq \psi(\xi_2).
\end{aligned}$$

□

When $A(y_1, \xi)$ satisfies the condition of the preceding lemma with $a = -\infty$ and $b = \xi_0$, the size of the two stage procedure is

$$\alpha = \sup_{\xi \leq \xi_0} \psi(\xi) = \psi(\xi_0). \tag{2.3.21}$$

The condition of Lemma 2.3.1 is difficult to check because there are infinitely many $A(y_1, \xi)$ for $\xi \in (-\infty, \xi_0)$. In addition to the Type I error rate, it is often of interest to specify the power at some alternative, ξ_1. When the power is specified to be ρ at $\xi = \xi_1$, it is desirable that the power is at least ρ for $\xi \in (\xi_1, \infty)$. Again, monotonicity of $\psi(\xi)$ is one way to guarantee that the power is at

least ρ for any ξ which is greater than ξ_1. To use Lemma 2.3.1 to guarantee the monotonicity of the power function in (ξ_1, ∞), we need to check $A(y_1, \xi)$ for all ξ in (ξ_1, ∞).

Because the condition of Lemma 2.3.1 involves infinitely many $A(y_1, \xi)$, we replace it with the following two conditions on the specification components of Stage II:

1. $A(y_1, \xi_1)$ is nondecreasing in $y_1 \in (k_1, k_2]$.

2. $n_2(y_1)$ is nonincreasing in $y_1 \in (k_1, k_2]$.

These two conditions only involve two specific functions and it is straightforward by direct computation to check these conditions. The motivation for imposing these conditions is two-fold. The first is that by Lemma 2.3.2 which follows, they guarantee that the power is monotone up to ξ_1 and it is at least $\psi(\xi_1)$ for $\xi > \xi_1$ though it may be decreasing for some range of values of $\xi \in (\xi_1, \infty)$. The second reason for imposing the conditions is a practical consideration. The larger the observed Stage I value, y_1, gets (in the interval $(k_1, k_2]$), the more evidence there is against H_0. Therefore, it is desirable to have the probability of rejecting H_0 under H_1 increase as y_1 increases, without a penalty of an increased Stage II sample size.

Lemma 2.3.2. *For an arbitrary ξ^*, if $A(y_1, \xi^*)$ is nondecreasingin y_1, and $n_2(y_1)$ is nonincreasing in y_1 for $y_1 \in (k_1, k_2]$, then,*

1. *$\psi(\xi)$ is nondecreasing in ξ for all $\xi \in (-\infty, \xi^*)$ and*

2. *$\psi(\xi) > \psi(\xi^*)$ for all $\xi \in (\xi^*, \infty)$.*

Proof: 1. Let $\xi^\dagger \in (-\infty, \xi^*)$. Using the connection among the specification components of Stage II, it can be shown that

$$z_{A(y_1,\xi^\dagger)} = z_{A(y_1,\xi^*)} + \sqrt{n_2(y_1)}\,(\xi^* - \xi^\dagger).$$

Then, because $z_{A(y_1,\xi^*)}$ and $n_2(y_1)$ are both nonincreasing in y_1 and $\xi^* - \xi^\dagger > 0$, we have that $z_{A(y_1,\xi^\dagger)}$ is nonincreasing. Consequently, $A(y_1, \xi^\dagger)$ for $\xi^\dagger \in (-\infty, \xi^*)$ is nondecreasing. So we can apply Lemma 2.3.1 to have the first result of the lemma.

2. Let $\xi^\dagger \in (\xi^*, \infty)$. Then by the definition of $A(y_1, \xi)$ in (2.2.6), $A(y_1, \xi^\dagger) > A(y_1, \xi^*)$. Thus,

$$\psi(\xi^\dagger) = \int_{-\infty}^{\infty} A(y_1, \xi^\dagger) g_{\xi^\dagger}(y_1) dy_1 > \int_{-\infty}^{\infty} A(y_1, \xi^*) g_{\xi^\dagger}(y_1) dy_1.$$

Since $A(y_1, \xi^*)$ is nondecreasing, we can apply the same reasoning as in the proof of Lemma 2.3.1 to get

$$\begin{aligned}\int_{-\infty}^{\infty} A(y_1, \xi^*) g_{\xi^\dagger}(y_1) dy_1 &= \int_{-\infty}^{\infty} A(y_1, \xi^*) \frac{g_{\xi^\dagger}(y_1)}{g_{\xi^*}(y_1)}\, g_{\xi^*}(y_1) dy_1 \\ &> \int_{-\infty}^{\infty} A(y_1, \xi^*) g_{\xi^*}(y_1) dy_1 = \psi(\xi^*),\end{aligned}$$

giving the second result of the lemma. □

When the conditions of Lemma 2.3.2 are met with $\xi^* = \xi_1$, we have a desirable property of the power function that the overall power is nondecreasing up to ξ_1 at which it is set to be $1-\beta$. And the power function is guaranteed not to go below $1-\beta$ when the true ξ exceeds ξ_1. Furthermore, from Lemma 2.3.2 (i), the Type I error rate is as described by (2.3.21). The conditions in Lemma 2.3.2 are sufficient and not necessary. The power function may be nondecreasing even though the noted conditions are not satisfied.

We provide, as an example, a two-stage adaptive design of which $A(y_1, \xi_0)$ and $n_2(y_1)$ do not meet the conditions of Lemma 2.3.2 and the power function is not nondecreasing. In this design, although the Type I error rate is set at Δ_0 to be $\alpha = .025$, the probability of rejecting H_0 is higher for some values of Δ that smaller than Δ_0. Consider testing

$$H_0 : \mu_t - \mu_c \leq \Delta_0$$
$$H_1 : \mu_t - \mu_c > \Delta_0,$$

where $\Delta_0 = 0$. Suppose that the desired power is $\rho = .9$ at $\Delta_1 = 1$. We assume that σ is known to be 4. Then the single-stage conventional design's sample size is $N = 337$. We choose the Stage I sample size to be 80% of N, i.e., $n_1 = 337 \times .8 = 270$. For the remaining two Stage I specification components, we choose $k_1 = -1.0$ and $\alpha_1 = .005$. With these three components specified, we can obtain the remaining components, β_1 and k_2. The Stage I is characterized by the following 5 specification components. $\{\alpha_1 = .005, \beta_1 = .000, n_1 = 270, k_1 = -1.0, k_2 = .1568\}$.

For Stage II specification components, we choose $A(y_1, \xi_0)$ and $A(y_1, \xi_1)$ that are not increasing in y_1. Let

$$A(y_1, \xi_0) = \begin{cases} .90 & \text{for } k_1 \leq y_1 < -.5 \\ -.20375 - 2.2075\, y_1 & \text{for } -.5 \leq y_1 < -.1 \\ .0201 & \text{for } -.1 \leq y_1 < k_2. \end{cases}$$

And let

$$A(y_1, \xi_1) = \begin{cases} .95 & \text{for } k_1 \leq y_1 < -.5 \\ .675 - .55\, y_1 & \text{for } -.5 \leq y_1 < -.1 \\ .73 & \text{for } -.1 \leq y_1 < k_2. \end{cases}$$

These A functions satisfy (2.2.7) and (2.2.8), respectively, so that the probability of rejecting H_0 is .025 and .90 at $\Delta = \Delta_0$ and $\Delta = \Delta_1$, respectively. See Figure 2.1 for these A-functions. Figure 2.2 shows the total sample size, $n_t(y_1) = n_1 + n_2(y_1)$ for this design. The Stage II sample size, $n_2(y_1)$, is not nonincreasing. So neither of the conditions of Lemma 2.3.2 is satisfied.

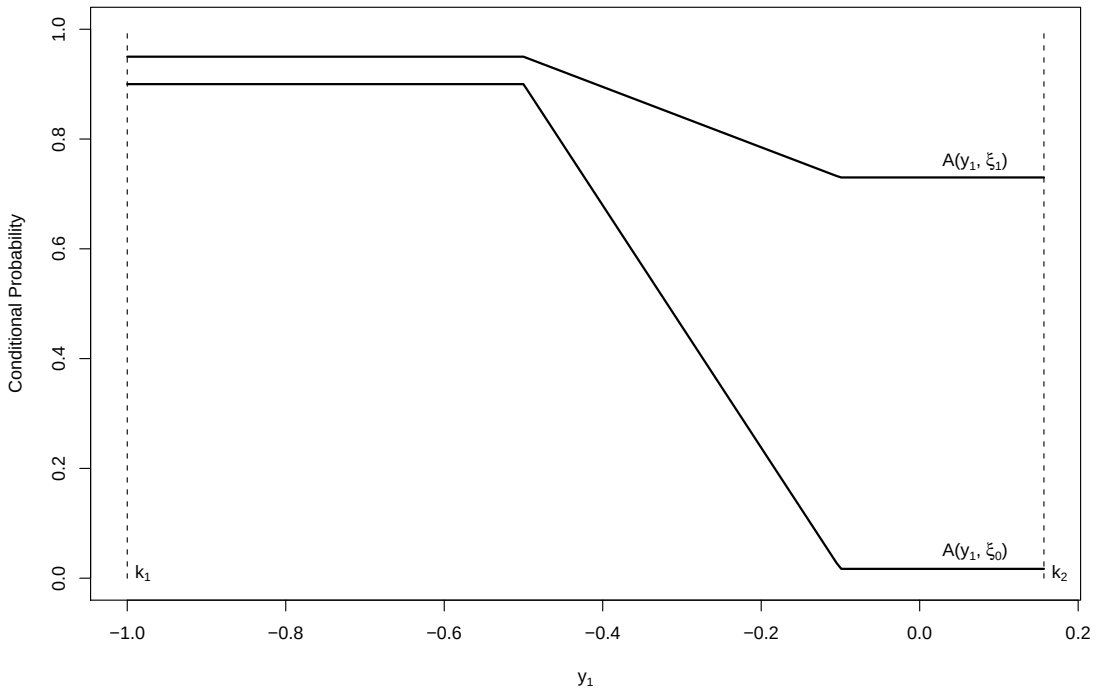


Figure 2.1: *Example of Section 2.3: The conditional power functions,* $A(y_1, \xi_0)$ *and* $A(y_1, \xi_1)$.

As can be seen in Figure 2.3, the power function of this procedure is not monotone. Although the probability of rejecting H_0 is set to be .025 at $\Delta = \Delta_0$, it is greater than .025 for a wide range of Δ's in (-∞, Δ_0). The probability of rejecting H_0 is as high as .511 in $(-\infty, \Delta_0)$. Therefore, obviously, the Type I error rate is not .025 for this design, even though the probability of rejecting H_0 is controlled at $\Delta = \Delta_0$. Table 2.2 summarizes the Stage I probabilities, power and expected sample sizes for this procedure.

As noted before, the conditions of Lemma 2.3.2 are sufficient ones and not necessary ones. Many designs that do not satisfy these conditions still have monotone power functions. However, as this example shows, it is possible for a two-stage adaptive procedure to have a power function that is not monotone. Therefore, for the two-stage adaptive procedures that we consider, we impose the two conditions of Lemma 2.3.2.

The conditions of Lemma 2.3.2 concern $A(y_1, \xi^*)$ for some ξ^* and $n_2(y_1)$. However, the two prespecified components for Stage II need not necessarily be $A(y_1, \xi^*)$ and $n_2(y_1)$. If not, $A(y_1, \xi^*)$ and $n_2(y_1)$ must be obtained through the functional relationships described in (2.2.12) and (2.2.13). It is then difficult to guarantee that $A(y_1, \xi^*)$ and $n_2(y_1)$ satisfy the conditions of Lemma 2.3.2. For example, specifying $A(y_1, \xi_0)$ and $A(y_1, \xi_1)$ such that $n_2(y_1)$ is nonincreasing for all $y_1 \in (k_1, k_2]$ is not a simple task. Or specifying $A(y_1, \xi_0)$ and $n_2(y_1)$ such that $A(y_1, \xi_1)$ is nondecreasing for

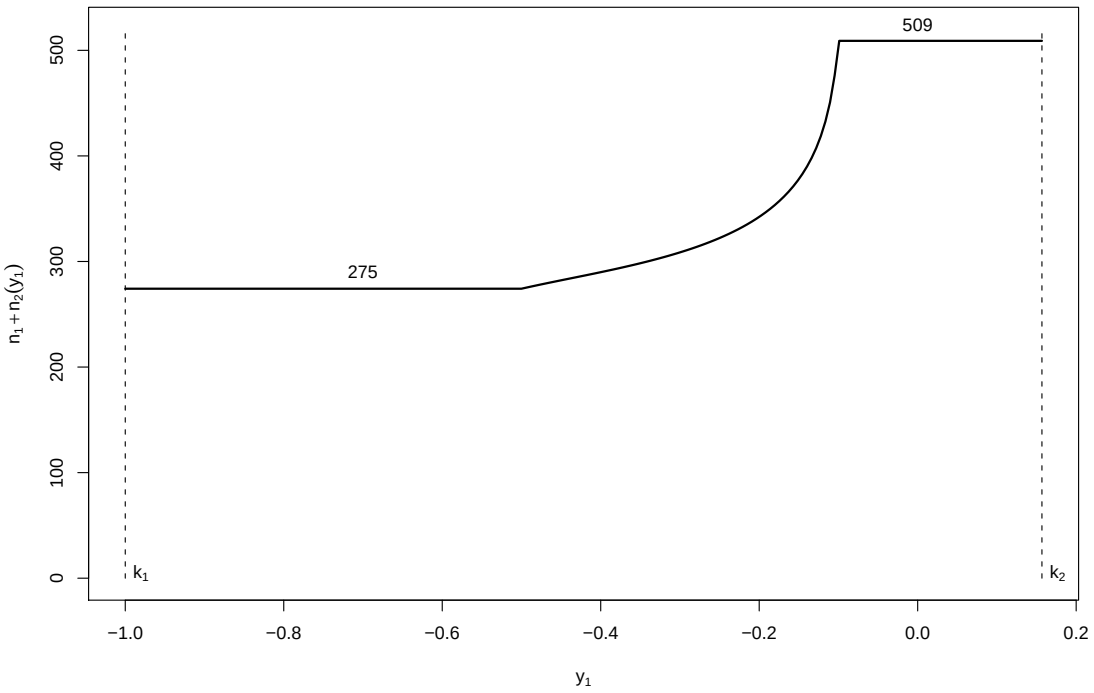


Figure 2.2: *Example of Section 2.3:* $n_1 + n_2(y_1)$.

all $y_1 \in (k_1, k_2]$ is also not a simple task. In such cases, the flexibility of our framework plays a significant role. The key fact is that it is relatively easy to make adjustments to the specifying components. For example, suppose that we specify $A(y_1, \xi_0)$ and $A(y_1, \xi_1)$ at the beginning, and the resulting $n_2(y_1)$ is not nonincreasing in $y_1 \in (k_1, k_2]$. Instead it is a convex function of y_1 in this range. Then we can modify the $n_2(y_1)$ function so that it is nondecreasing in y_1 while keeping $A(y_1, \xi_0)$ fixed. When $A(y_1, \xi_0)$ is fixed and $n_2(y_1)$ is altered, the other two components, $A(y_1, \xi_1)$ and $w(y_1)$ are reobtained accordingly. Thus we have a new set of $A(y_1, \xi_0)$, $A(y_1, \xi_1)$, $n_2(y_1)$ and $w(y_1)$. It is, of course, desirable to make the new $A(y_1, \xi_1)$ satisfy the condition (2.2.8). Even when the conditions of Lemma 2.3.2 are satisfied, if there still is need to modify the design to satisfy certain design characteristics (e.g., $A(y_1, \xi_1)$ is too small for some $y_1 \in (k_1, k_2]$) then we can continue altering the components until the design is satisfactory. We use the following examples to demonstrate the flexibility of our framework for designing two-stage adaptive trials.

2.4 Examples

In this section, we use two examples to demonstrate the flexibility of our procedure in constructing a two-stage adaptive design. First, we show how a two-stage group sequential method can be placed into our framework. Using tabulated critical values for Stage I and for Stage II from Jennison

Table 2.2: *Example of Section 2.3: Stage I probabilities, power and expected sample sizes.*

Δ	Stage I			Stage II	Power	$E[n_t(y_1)]$
	Accept	Continue	Reject	Reject		
-4.00	.000	1.000	.000	.432	.432	274.2
-3.00	.000	1.000	.000	.480	.480	276.1
-2.00	.000	1.000	.000	.094	.094	299.1
-1.00	.000	1.000	.000	.018	.018	375.5
0.00	.000	.995	.005	.020	.025	503.7
1.00	.000	.371	.629	.271	.900	358.8

and Turnbull (1999), we find the corresponding A-functions. Our calculation of the Type I error rate and the power agrees with their result.

In the second example, we consider a procedure proposed by Lan and Trost (1996). Their procedure is atypical because Stage II is divided into two pieces based on y_1, and each part has different specification components. However, by treating these two parts of Stage II separately, we can place their method into our framework. In their paper, they obtain the Type I error rate using simulations, but with our method, it is computed analytically using numerical integration. Their maximum Stage II sample size in this particular example is 3222 whereas N is only 288. The modification of this procedure using our framework yields the same Type I error rate, the same power at the original alternative while allowing the maximum sample size to be prespecified.

Both examples demonstrate our framework's flexibility and show that the framework can encompass a variety of two-stage adaptive design. The second example, in addition, shows the framework's ability to facilitate a two-stage adaptive design with specific characteristics in terms of Type I error rate, power and expected and maximum sample size.

2.4.1 Two-Stage Group Sequential Procedure

Our framework encompasses the conventional one-sided two-stage group sequential tests. The two-stage conventional group sequential procedure is more restricted than what our framework is capable of because the Stage II sample size, $n_2(y_1)$ and the Stage II critical value, $c(y_1)$ for the group sequential method do not vary with y_1.

Jennison and Turnbull (1999) give a discussion of a general one-sided group sequential test. They provide sample sizes and critical values for Stage I and Stage II. Their discussion of the procedure is in terms of the standardized Z statistics. To comply with their definition, let

$$Z_1 = \sqrt{n_1}\, Y_1, \tag{2.4.22}$$

$$Z_b = \sqrt{n_1 + n_2}\, Y_b, \tag{2.4.23}$$

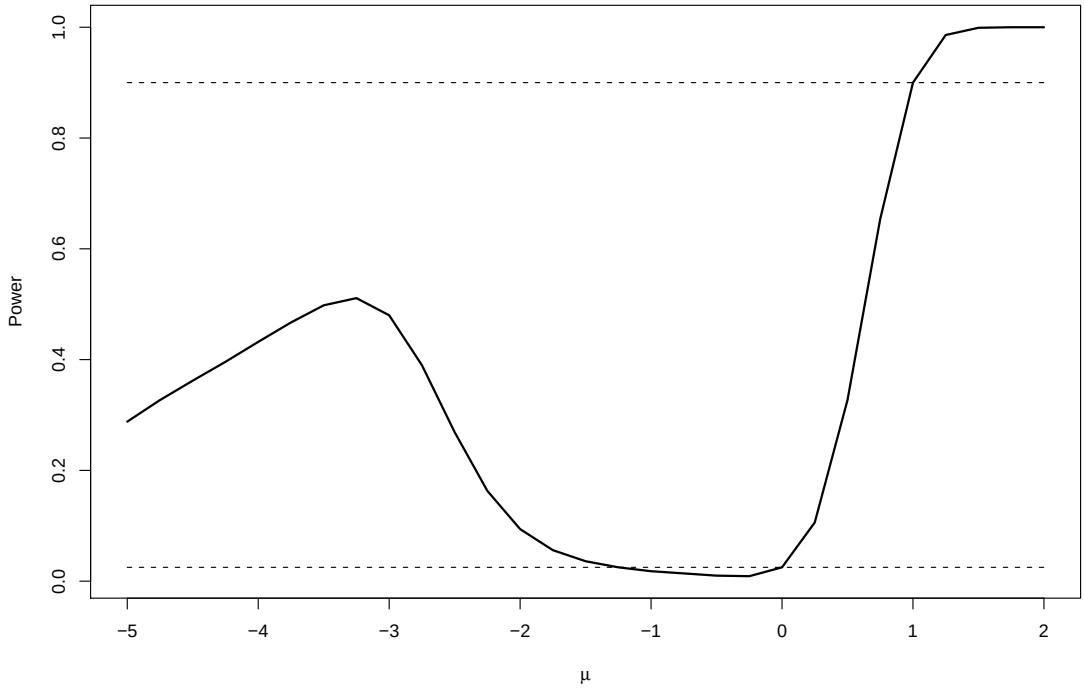


Figure 2.3: *Example of Section 2.3: The power function.*

where Y_1 is defined in (2.2.1) and Y_b is given by (2.2.4) but with n_2 that does not depend on y_1. A general two-stage one-sided group sequential test is defined by three constants, a_1, b_1 and a_2, where $a_1 < b_1$. A decision is made at the end of Stage I according to the following decision rule, which is based on the standardized Stage I statistic, Z_1. At the end of Stage I,

$$\text{if} \begin{cases} Z_1 \leq a_1 & \text{stop and accept } H_0, \\ a_1 < Z_1 \leq b_1 & \text{continue to Stage II}, \\ b_1 < Z_1 & \text{stop and reject } H_0. \end{cases}$$

The decision rule at the end of Stage II is written similarly in terms of the standardized statistic, Z_b, from the combined data. At the end of Stage II,

$$\text{if} \begin{cases} Z_b \leq a_2 & \text{stop and accept } H_0, \\ a_2 < Z_b & \text{stop and reject } H_0. \end{cases}$$

In order to place this group sequential method into our framework, the following two characteristics of the group sequential method are noticed.

1. The Stage II sample size, n_2, is the same for all y_1.

2. The Stage II critical value, c, for Y_b is the same for all y_1.

Jennison and Turnbull (1999) only give the constants $\tilde{C}_1$ and $\tilde{C}_2$ that are used to calculate critical values in the case with an equally spaced information level, meaning that n_1 and n_2 are equal.

However, the case where n_2 is not the same as n_1 is also easily implemented in our framework. Let $n_2 = f^* n_1$, where f^* is any positive number. Now since both c and n_2 do not depend on y_1, we can write $w(y_1)$, using (2.2.5), as follows:

$$w(y_1) = \frac{n_1}{n_2}(c - y_1) + c = (1 + \frac{1}{f^*})c - \frac{y_1}{f^*} \tag{2.4.24}$$

Substituting (2.4.24) in (2.2.16), we get

$$A(y_1, \xi) = 1 - \Phi\left[\sqrt{f^* n_1}\left((1 + \frac{1}{f^*})\, c - \frac{y_1}{f^*} - \xi\right)\right] \tag{2.4.25}$$

Note that if $f^* = 1$, that is, if the Stage II sample size is equal to the Stage I sample size, then (2.4.24) and (2.4.25) simplify to

$$\begin{aligned} w(y_1) &= 2c - y_1, \\ A(y_1, \xi) &= 1 - \Phi\left[\sqrt{n_1}\,(2c - y_1 - \xi)\right]. \end{aligned} \tag{2.4.26}$$

In this example, we use Jennison and Turnbull's tabulated values to find the critical values and sample sizes for Stage I and II, and obtain the corresponding A-functions. We show, by calculating the Type I error rate and the power, that our result agrees to theirs. Consider the following situation. The hypotheses of interest are $H_0 : \mu_t - \mu_c \leq 0$ against $H_1 : \mu_t - \mu_c > 0$. The error probabilities are $\alpha = .05$ and $\beta = .1$ at $\Delta = \Delta_1 = 1$. We assume that $\sigma = 4$. Note that $\xi_0 = 0$ and $\xi_1 = \Delta_1/(\sqrt{2}\,\sigma) = .1768$.

From Table 4.2 in Jennison and Turnbull (1999), we have $\tilde{C}_1 = 1.657$ and $\tilde{C}_2 = 1.332$ for $K = 2$ (two stages). The total sample size, n_b, and the Stage I sample size, n_1, are

$$\begin{aligned} n_b &= \frac{(\tilde{C}_1 + \tilde{C}_2)^2}{\xi_1^2} = \frac{(1.657 + 1.332)^2}{.1768^2} = 285.9 \approx 286, \\ n_1 &= (1/2) n_b = 143. \end{aligned}$$

The Stage I critical values, a_1 and b_1, for the standardized test statistic are given by

$$\begin{aligned} a_1 &= \xi_1 \sqrt{n_1} - \tilde{C}_2 (1/2)^{-1/2} = .1768\sqrt{142.9} - 1.332(1/2)^{-1/2} = .2298, \\ b_1 &= \tilde{C}_1 (1/2)^{-1/2} = 1.657(1/2)^{-1/2} = 2.343. \end{aligned}$$

The Stage II critical value for the standardized test statistic is

$$a_2 = .25\sqrt{142.9} - 1.332 = 1.657.$$

Using (2.4.22) and (2.4.23) we can transform these critical values to k_1, k_2 and c to be used with Y_1 and Y_b. We have

$$\begin{aligned} k_1 &= a_1/\sqrt{n_1} = .2298/\sqrt{143} = .0192, \\ k_2 &= b_1/\sqrt{n_1} = 2.343/\sqrt{143} = .1959, \\ c &= a_2/\sqrt{n_1 + n_2} = 1.657/\sqrt{286} = .0980. \end{aligned}$$

With these k_1 and k_2, we can calculate the Stage I probabilities of accepting H_0, of continuing to Stage II and of rejecting H_0 under H_0 and H_1. These are tabulated in Table 2.3.

Using $f^* = 1$, and thus the A-functions in (2.4.26), we get

$$A(y_1, \xi_0) = 1 - \Phi\left[\sqrt{143}\,(2 \times .0980 - y_1)\right] = 1 - \Phi\left[2.34 - 11.96 y_1\right],$$
$$A(y_1, \xi_1) = 1 - \Phi\left[\sqrt{143}\,(2 \times .0980 - y_1 - .1768)\right] = 1 - \Phi\left[.230 - 11.96 y_1\right].$$

These two A-functions are displayed in Figure 2.4. Note that both $A(y_1, \xi_0)$ and $A(y_1, \xi_1)$ are nondecreasing in y_1. The resulting total sample size when continuing to Stage II is uniformly 143 as planned. Therefore, by Lemma 2.3.2, the power function, $\psi(\xi)$ is nondecreasing in ξ. As Table 2.3 shows, the Type I error rate of this design is $\alpha = .05$ and the power at the original alternative is $1 - \beta = .90$. The expected sample size is reduced compared to the sample size, $N = 274$, of the conventional single-stage design. The percentage reduction in expected sample size from N is 27% when $\xi = \xi_0$, 19% when $\xi = \xi_1$ and 12% when $\xi = \xi_1/2$. These results agree with Jennison and

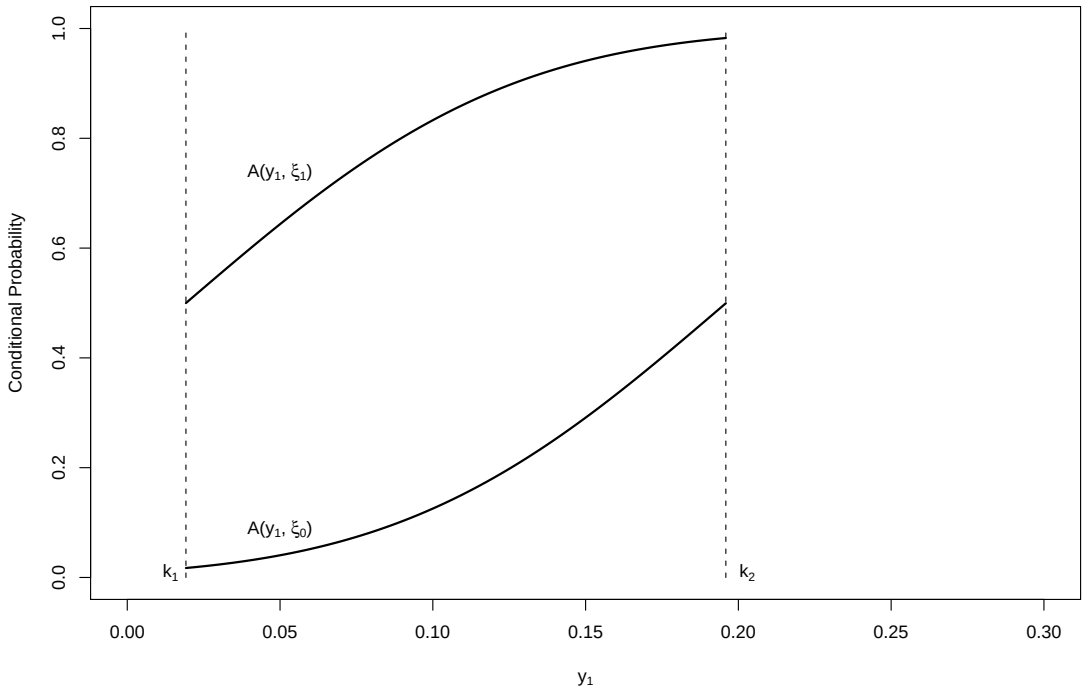


Figure 2.4: *Example of Section 2.4.1: $A(y_1, \xi_0)$ and $A(y_1, \xi_1)$ for a group sequential procedure.*

Turnbull's calculations.

Table 2.3: *Example of Section 2.4.1: Group Sequential Procedure. Stage I probabilities, power and expected sample sizes.*

Δ	Stage I			Stage II	Power	$E[n_t(y_1)]$
	Accept	Continue	Reject	Reject		
0.00	.591	.400	.009	.041	.050	199.2
0.25	.383	.582	.035	.146	.181	225.2
0.50	.205	.695	.100	.329	.429	241.4
0.75	.088	.686	.225	.485	.710	240.2
1.00	.030	.560	.410	.490	.900	222.1
1.25	.008	.374	.617	.360	.977	195.6

2.4.2 Lan and Trost's Procedure And Its Modifications

In this example, we consider a procedure proposed by Lan and Trost (1997). Their Stage I sample size is expressed as a fraction of N. At the end of Stage I, they calculate CP, the conditional power under the Stage I result and apply the following rules:

1. If $CP \leq c_l$ (lower limit), then stop the trial and accept H_0.
2. If $CP \geq c_u$ (upper limit), then continue to the scheduled end of the study and reject H_0 if the final $Z \geq z_\alpha$.
3. If $c_l < CP < c_u$, then extend the study to a sample size of mN $(m > 1)$ so that CP_m, the conditional power under the current trend after extension, is c_u. Reject H_0 if Z_m, the final Z-statistic after extension, exceeds z_α.

Conceptually, CP is a function of y_1, so we denote it by $B(y_1)$. Then, k_1 satisfies $B(k_1) = c_l$. Let κ be the point such that $B(\kappa) = c_u$. For this procedure, the Stage I specification components are n_1, k_1 and k_2. Note that $k_2 = \infty$ because the trial does not stop at the end of Stage II to reject H_0, thus resulting in $\alpha_1 = 0$. The specification of Stage II components is divided into two regions based on y_1. The first region is where $y_1 \in (k_1, \kappa)$ and the second is where $y_1 \in (\kappa, \infty)$. First we show how to obtain k_1 and κ.

The decision rule at the end of Stage II is $Z > z_\alpha$. Since $Y_b = Z/\sqrt{n_1 + n_2}$, it can be written in terms of Y_b as $Y_b > z_\alpha/\sqrt{n_1 + n_2}$. Using (2.2.5), we get

$$w(y_1) = \frac{\sqrt{n_1 + n_2(y_1)}}{n_2(y_1)} z_\alpha - \frac{n_1}{n_2(y_1)} y_1. \tag{2.4.27}$$

Now we can express $B(y_1)$ using $n_2(y_1)$ and N. We use the critical value, $w(y_1)$, in (2.4.27) but with a constant Stage II sample size $n_2(y_1) = N - n_1$ to find:

$$\begin{aligned} B(y_1) =& 1 - \Phi\left[\sqrt{N-n_1}\,(w(y_1) - y_1)\right] \\ =& 1 - \Phi\left[\frac{\sqrt{N}}{\sqrt{N-n_1}}\,z_\alpha - \frac{N}{\sqrt{N-n_1}}\,y_1\right]. \end{aligned}$$

Then given N and n_1, we can solve $B(k_1) = c_l$ and $B(\kappa) = c_u$ to obtain

$$k_1 = \frac{1}{\sqrt{N}}\,z_\alpha - \frac{\sqrt{N-n_1}}{N}\,z_{c_l}, \tag{2.4.28}$$

$$\kappa = \frac{1}{\sqrt{N}}\,z_\alpha - \frac{\sqrt{N-n_1}}{N}\,z_{c_u}. \tag{2.4.29}$$

Now we shift our attention to Stage II specification components. For the region $y_1 \in (k_1, \kappa)$, $A(y_1, y_1) = c_u$ for all y_1. So we have

$$A(y_1, y_1) = c_u = \Phi\left[\sqrt{n_2(y_1)}(w(y_1) - y_1)\right]. \tag{2.4.30}$$

We can simultaneously solve (2.4.27) and (2.4.30) numerically for $w(y_1)$ and $n_2(y_1)$. Then using these $w(y_1)$ and $n_2(y_1)$, we find the conditional probability of rejecting H_0 under any ξ by (2.2.16).

For the region $y_1 \in (\kappa, \infty)$, we have a fixed $n_2 = N - n_1$ and $A(y_1, y_1) = B(y_1)$. Using these two elements, we can find

$$\begin{aligned} A(y_1, \xi) =& 1 - \Phi\left[z_{B(y_1)} + \sqrt{N-n_1}\,(y_1 - \xi)\right], \\ w(y_1) =& \frac{z_{B(y_1)}}{\sqrt{N-n_1}} + y_1. \end{aligned}$$

Consider a specific example. Let $\Delta_0 = 0$, $\Delta_1 = 1$ and $\sigma = 4$. Then $\xi_0 = 0$ and $\xi_1 = .1768$. With $\alpha = .025$ and $\beta = .15$, we obtain $N = 288$. Further suppose that $f = n_1/N = .4$. Then $n_1 = 115$ and $n_2 = 173$. Let c_l be .05 and c_u be .65. Then using (2.4.28) and (2.4.29) we get $k_1 = .0405$ and $\kappa = .1332$.

The conditional power functions are displayed in Figure 2.5. The dotted curve line represents $B(y_1)$. The three solid lines are conditional power functions. The bottom one is $A(y_1, \xi_0)$, the middle one is $A(y_1, y_1)$, and the top one is $A(y_1, \xi_1)$. Note that $A(y_1, y_1)$ and $A(y_1, \xi_1)$ cross each other at $y_1 = \xi_1$. Also note that $A(y_1, \xi_1)$ is not monotonically increasing. Figure 2.6 shows $n_b(y_1)$ (the total sample size) of this procedure. The maximum of $n_b(y_1)$ occurs at $y_1 = k_1$ and it is 3338. ($n_2(k_1) = 3338 - 115 = 3222$).

Because $A(y_1, \xi_0)$ is nondecreasing in y_1 and $n_2(y_1)$ is nonincreasing in y_1, Lemma 2.3.2 can be applied with $\xi^* = \xi_0$ to guarantee that Type I error rate is $\int_{k_1}^{k_2} A(y_1, \xi_0) g_{\xi_0}(y_1)\, dy_1$. However,

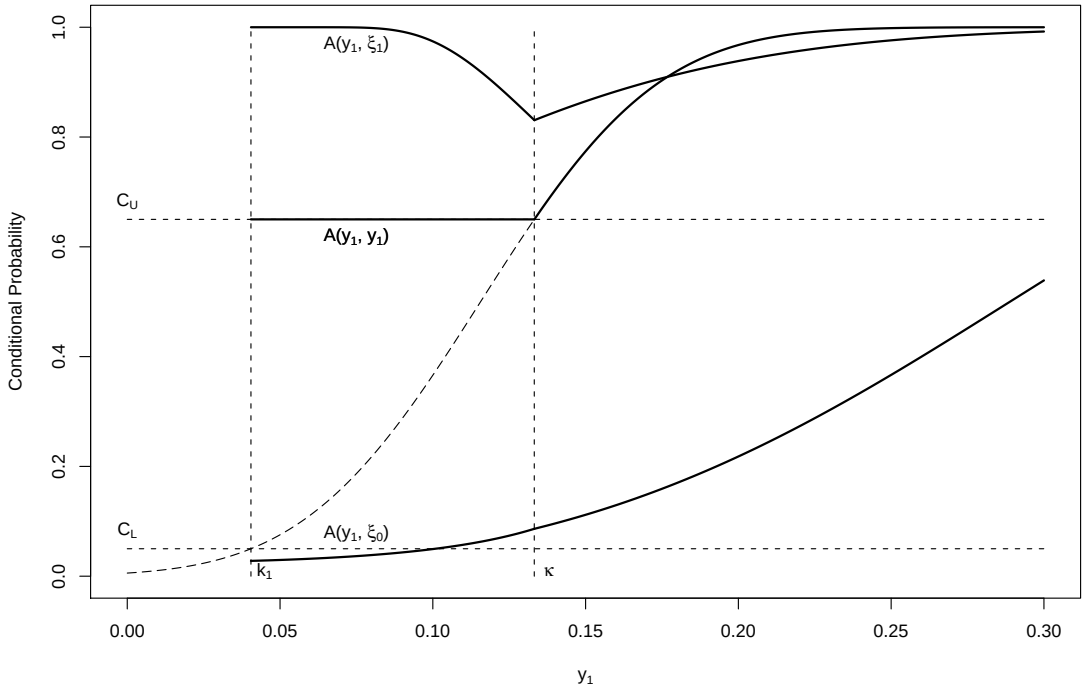


Figure 2.5: *Example of Section 2.4.2: $A(y_1, \xi_0)$ and $A(y_1, \xi_1)$ for Lan and Trost's original procedure.*

note that $A(y_1, \xi_1)$ is not monotonically increasing. Therefore, monotonicity of the power function in $y_1 \in (-\infty, \xi_1)$ cannot be guaranteed by Lemma 2.3.2. However, numerical computation does show that the power function is increasing in $\xi \in (-\infty, \infty)$.

In this example, although $A(y_1, \xi_1)$ is nonincreasing, the power function is increasing for all $\xi \in (-\infty, \infty)$. However, from a practical viewpoint, this $A(y_1, \xi_1)$ gives the following contradiction. As shown in Figure 2.5, the conditional probability of rejecting H_0 under H_1 is virtually 1 if we observe $y_1 = .0405$, which is on the borderline of accepting H_0 and continuing to Stage II. On the other hand, if we observe $y_1 = .1332$, then the conditional power is down to about .8 even though there is stronger evidence against H_0 in Stage I. Also, as seen in this example, forcing $A(y_1, y_1)$ to be fairly high (in this case .65) is not practical because it inflates the necessary sample size when y_1 is close to ξ_0. If $k_1 \leq \xi_0$ (which is possible in their formulation), $n_2(y_1) = \infty$ at $y_1 = \xi_0$.

Since we have $A(y_1, \xi_0)$ and $A(y_1, \xi_1)$, we can calculate for the Lan and Trost procedure the Type I error rate and the unconditional power at the original alternative, ξ_1, using (2.3.20). The type I error rate is .024, which agrees with their simulation result. The power at ξ_1 is .877.

We now show how to use our technique to modify the Lan and Trost design to effect a substantial improvement. In order to make the new design comparable to the original Lan and Trost design, we set $\alpha = .024$ and the power at ξ_1 equal to .877; the corresponding conventional one-stage sample

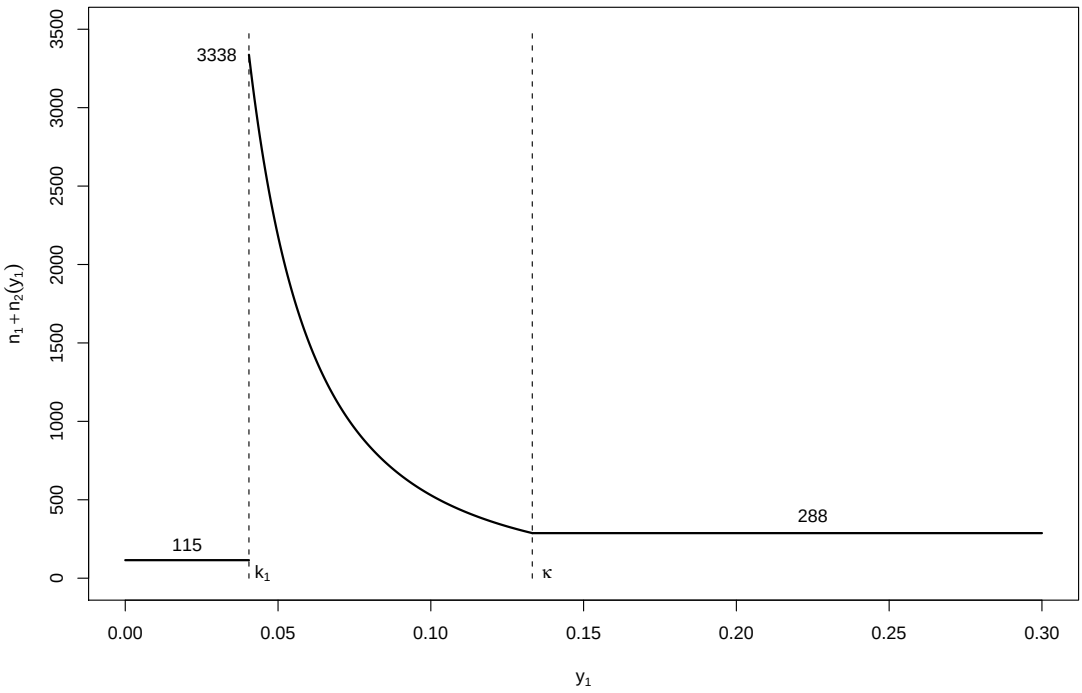


Figure 2.6: *Example of Section 2.4.2: Total sample size for Lan and Trost's original procedure.*

size is $N = 316$. The maximum total sample size is set at 442, a 40% increase from N. We use the same $n_1 = 115$. Further, k_1 is kept fixed at .0405, so that β_1 remains .072. We change k_2 from ∞ to .2406, so that $\alpha_1 = .005$. Hence $\alpha_2 = .019$ and $\beta_2 = .051$.

For Stage II specification components, first we choose $A(y_1, \xi_0)$ to have a quadratic form. $A(y_1, \xi_0) = .02 + 6.93(y_1 - k_1)^2$. This function satisfies the condition, $\alpha_2 = \int_{k_1}^{k_2} A(y_1, \xi_0) g_0(y_1)\, dy_1$, and it is very similar to the Lan and Trost's original $A(y_1, \xi_0)$ function over values of y_1 in (k_1, k_2).

For $A(y_1, \xi_1)$, we also choose a quadratic function: $A(y_1, \xi_1) = .89 + 2.4(y_1 - k_1)^2$. Using these $A(y_1, \xi_0)$ and $A(y_1, \xi_1)$, we can calculate $n_2(y_1)$ and $w(y_1)$. For $n_2(y_1)$, in addition to the noted condition that $\max(n_2(y_1)) = 442 - n_1 = 327$, we impose a condition that $n_2(y_1) \geq 201$ which makes the minimum value of $n_1 + n_2(y_1)$ equal to $N = 316$. These two conditions provide a flat $n_2(y_1) = 327$ for y_1 in $(k_1, .0685)$ and a flat $n_2(y_1) = 201$ for y_1 in $(.202, k_2]$. Keeping $A(y_1, \xi_0)$ unchanged and modifying $n_2(y_1)$ to accommodate these maximum and minimum requirements alters $A(y_1, \xi_1)$. The final $A(y_1, \xi_1)$ is:

$$A(y_1, \xi_1) = \begin{cases} 1 - \Phi\left[z_{A(y_1,\xi_0)} + \sqrt{327}\,(.1768)\right] & \text{if } k_1 < y_1 < .0685 \\ .89 + 2.4(y_1 - k_1)^2 & \text{if } .0685 \leq y_1 < .202 \\ 1 - \Phi\left[z_{A(y_1,\xi_0)} + \sqrt{201}\,(.1768)\right] & \text{if } .202 \leq y_1 < k_2 \end{cases}$$

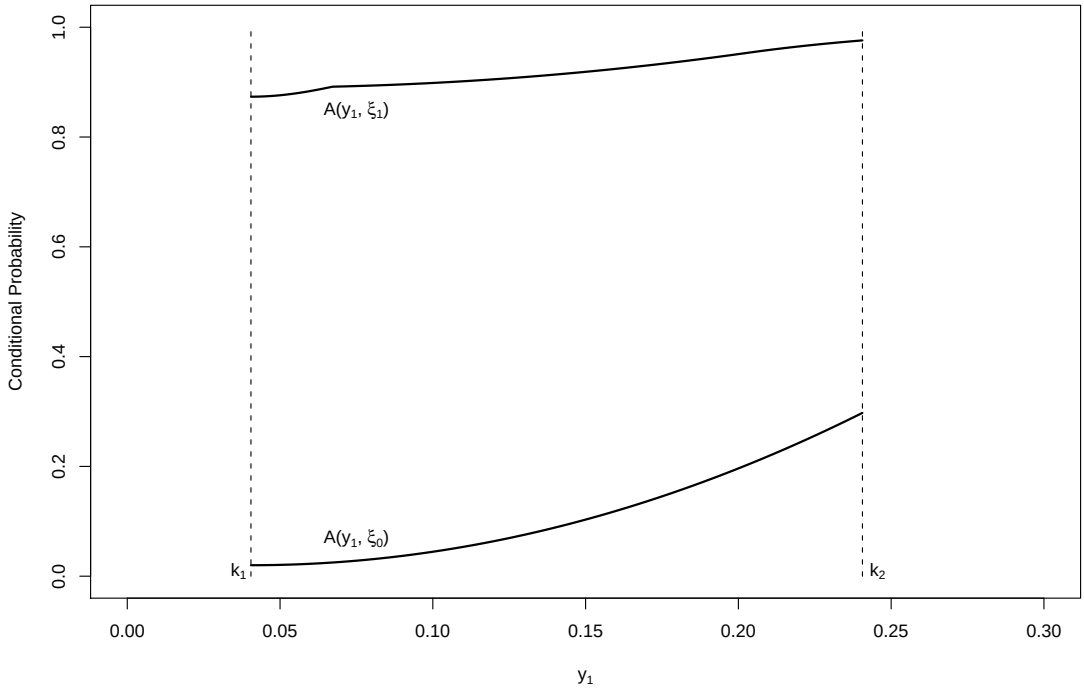


Figure 2.7: *Example of Section 2.4.2: $A(y_1, \xi_0)$ and $A(y_1, \xi_1)$ for the modified procedure.*

Figure 2.7 shows the conditional power functions. The total sample size is shown in Figure 2.8. Tables 2.4 and 2.5 summarize the power and the expected sample size for various values of Δ in $[0, 1.25]$ for the original Lan and Trost's design and for the modified design. While the powers at $\Delta = 0$ and at $\Delta = 1$ are the same (by design), the powers at intermediate points are somewhat higher in the Lan and Trost design. However, from our perspective, the most drastic differences are in expected sample sizes. The percentage reduction in expected sample size is 48% when $\Delta = 0$, 43% when $\Delta = .5$ and 31% when $\Delta = 1$. Furthermore, the maximum total sample size for the modified design is 442, whereas it is 3338 for the original design.

Table 2.4: *Example of Section 2.4.2: Lan & Trost's original design. Stage I probabilities, power and expected sample sizes.*

Δ	Stage I			Stage II	Power	$E[n_t(y_1)]$
	Accept	Continue	Reject	Reject		
0.00	.668	.332	0	.024	.024	282.4
0.25	.484	.516	0	.178	.178	361.6
0.50	.304	.696	0	.457	.457	386.0
0.75	.162	.838	0	.706	.706	353.7
1.00	.072	.928	0	.877	.877	292.7
1.25	.026	.974	0	.961	.961	235.8

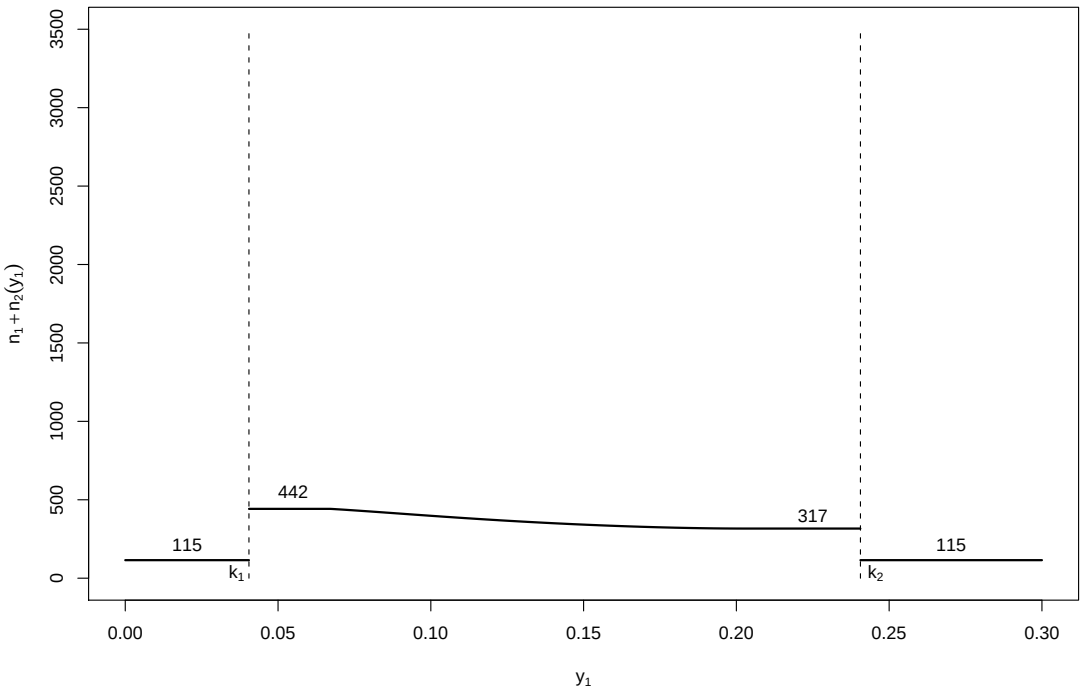


Figure 2.8: *Example of Section 2.4.2: Total sample size for the modified procedure.*

2.5 Design Parameters And Their Relative Significance

To construct a two-stage adaptive design, the various design components have to be specified. For example, in Stage I, one can choose α_1, β_1 and n_1 and for Stage II, any $A(y_1, \xi_0)$ and $A(y_1, \xi_1)$ can be used as long as they satisfy the conditions (2.2.7) and (2.2.8). In this section, we discuss the results of a computational experiment that examines which of these components have significant influence on the characteristics of the resulting design.

Different combinations of the values of design components of interest yield different designs with different performance characteristics. For each performance characteristic, we investigate its "response surface" as a function of the design factor combinations. A design component that demonstrates relatively substantial variability in the "response surface" is considered influential; if the "response surface" does not vary much for different values of a particular component, then the component is considered unimportant in determining performance characteristics.

More specifically, consider the setting where $H_0 : \mu_t - \mu_c \leq 0$ is tested against $H_1 : \mu_t - \mu_c > 0$. Let Y_1 and Y_2 be defined, respectively, by (2.2.1) and (2.2.3). Let $\alpha = .05$ and $\beta = .10$ at the chosen alternative ξ_1, where ξ_1 varies among $\{.125, .2, .25\}$.

Table 2.5: *Example of Section 2.4.2: The modified design. Stage I probabilities, power and expected sample sizes.*

Δ	Stage I			Stage II	Power	$E[n_t(y_1)]$
	Accept	Continue	Reject	Reject		
0.00	.668	.327	.005	.020	.025	191.3
0.25	.484	.494	.018	.097	.115	226.6
0.50	.304	.644	.052	.280	.332	252.9
0.75	.162	.714	.124	.500	.624	260.8
1.00	.072	.680	.248	.602	.850	247.8
1.25	.026	.555	.418	.538	.956	219.4

Let the Stage I sample size, n_1, have the form $n_1 = f \times N$, where $0 < f < 1$, where f varies among {.3, .5, .75}, and N is the sample size needed for the conventional single-stage procedure to achieve the same α and the power. Given ξ_1 and f, n_1 can be calculated. Let α_1 and β_1, respectively, vary among {.01, .025, .04} and {.02, .05, .08}; so that for each combination of α_1, β_1, f and ξ_1, k_1 and k_2 are uniquely determined.

For Stage II, in order to see the effect of the conditional power functions, five choices for each of $A(y_1, \xi_0)$ and $A(y_1, \xi_1)$ are included in the study. The first choice is $A_{Flat}(y_1, \xi_0) = a_{0F}$, which takes a constant value, a_{0F}, for all $y_1 \in (k_1, k_2]$. For each combination of Stage I components, a_{0F} is determined so that the Type I error rate is α by solving

$$\alpha_2 = a_{0F} \int_{k_1}^{k_2} g_{\xi_0}(y_1)\, dy_1 = a_{0F}(\Phi\left[\sqrt{n_1}\, k_2\right] - \Phi\left[\sqrt{n_1}\, k_1\right]).$$

The second is $A_{LC}(y_1, \xi_0)$, which is a generalization by Liu and Chi (2001) of the $A(y_1, \xi_0)$ that Proschan and Hunsberger (1995) used. For $A_{LC}(y_1, \xi_0)$, we need to find γ that satisfies

$$\alpha_2 = \int_{k_1}^{k_2} \left(1 - \Phi\left[\gamma\sqrt{n_1}\sqrt{(k_2 - \xi_0)^2 - (y_1 - \xi_0)^2}\right]\right) g_{\xi_0}(y_1)\, dy_1.$$

The third A-function is a straight line with the form, $A_{Linear}(y_1, \xi_0) = a_0 + a_{1L}(y_1 - k_1)$. The fourth type is a step function, $A_{Step}(y_1, \xi_0) = a_0 + a_{1S} I(y_1 > (k_1 + k_2)/2)$. And finally, the fifth A-function is a square root function of the form, $A_{Root}(y_1, \xi_0) = a_0 + a_{1R}\sqrt{y_1 - k_1}$. Note that the last three A-functions all have the same value, a_0, at $y_1 = k_1$, where $0 \le a_0 \le a_{0F}$. To vary the $A-$functions and yet have some consistency, we require $A_{LC}(y_1, \xi_0)$, $A_{Linear}(y_1, \xi_0)$, $A_{Step}(y_1, \xi_0)$ and $A_{Root}(y_1, \xi_0)$ all have the same value, a_0, at $y_1 = k_1$. A concern is that without careful attention, we would inadvertently make these $A-$functions similar to each other, so that the effect of *different* $A-$functions could not be well assessed. For example, if $a_0 = a_{0F}$, $A_{Linear}(y_1, \xi_0)$, $A_{Step}(y_1, \xi_0)$ and $A_{Root}(y_1, \xi_0)$ would all be identical to $A_{Flat}(y_1, \xi_0)$. However, our preliminary findings suggest that $A_{LC}(y_1, \xi_0)$ at $y_1 = k_1$ is usually very small, and thus very different from a_{0F}, allowing the desired variability.

By fixing a_0, we can solve numerically for a_{1L} in $A_{Liner}(y_1, \xi_0)$, a_{1S} in $A_{Step}(y_1, \xi_0)$ and a_{1R} in $A_{Root}(y_1, \xi_0)$ so that the Type I error rate is α. For example, a_{1R} is computed by solving the following:

$$\begin{aligned}\alpha_2 &= \int_{k_1}^{k_2} (a_0 + a_{1R}\sqrt{y_1 - k_1}\,) g_{\xi_0}(y_1)\, dy_1 \\ &= a_0(\Phi\left[\sqrt{n_1}\, k_2\right] - \Phi\left[\sqrt{n_1}\, k_1\right]) + a_{1R} \int_{k_1}^{k_2} \sqrt{y_1 - k_1}\, g_{\xi_0}(y_1)\, dy_1.\end{aligned}$$

Similarly, five functions for $A(y_1, \xi_1)$ of the same forms as $A(y_1, \xi_0)$ are considered. They are

$$\begin{aligned} A_{Flat}(y_1, \xi_1) =& b_{0F}, \\ A_{LC}(y_1, \xi_1) =& 1 - \Phi\left[\gamma\sqrt{n_1}\,\sqrt{(\xi_1 - k_1)^2 - (\xi_1 - y_1)^2}\,\right], \\ A_{Linear}(y_1, \xi_1) =& b_{0L} + b_{1L}(y_1 - k_1), \\ A_{Step}(y_1, \xi_1) =& b_{0S} + b_{1S} I(y_1 > (k_1 + k_2)/2), \\ A_{Root}(y_1, \xi_1) =& b_{0R} + b_{1R}\sqrt{y_1 - k_1}\,. \end{aligned}$$

For $A_{Flat}(y_1, \xi_1)$, b_{0F} can be found analogously to a_{0F}. The form of $A_{LC}(y_1, \xi_1)$ is obtained through the same optimization that Proschan and Hunsberger (1995) used in deriving their $A(y_1, \xi_0)$ function. For the other three functions, b_0 and b_1 are calculated to satisfy

$$A_{LC}(k_2, \xi_1) = A_{Linear}(k_2, \xi_1) = A_{Step}(k_2, \xi_1) = A_{Root}(k_2, \xi_1).$$

In other words, they all have the same ending point (i.e., when $y_1 = k_2$) as $A_{LC}(y_1, \xi_1)$.

We observe that some combinations of $A(y_1, \xi_0)$ and $A(y_1, \xi_1)$ actually yield an increasing $n_2(y_1)$. We checked each such design that does not satisfy the condition (ii) of Lemma 2.3.2 and found for these that nonetheless the power functions are increasing.

These values of all the design components are summarized in Table 2.6. Note that there are 2025 possible combinations.

We now specify the three criteria used to compare designs' performance characteristics. The first is the power at three intermediate points, $\lambda\xi_1$ where $\lambda = \frac{1}{4}$, $\frac{1}{2}$ and $\frac{3}{4}$. By design the powers at $\lambda = 0$ and 1 are set to be .05 and .90, respectively. To compare the expected sample sizes of the procedures, we define the following two ratios:

$$\begin{aligned} &\text{(i) } R_1 = \frac{n_1 + E[n_2(Y_1)]}{N}\text{, at } \lambda\xi_1 \text{ where } \lambda = 0,\ \tfrac{1}{4},\ \tfrac{1}{2},\ \tfrac{3}{4} \text{ and } 1. \\ &\text{(ii) } R_2 = \frac{n_1 + E[n_2(Y_1)]}{N^*}\text{, at } \lambda\xi_1 \text{ where } \lambda = \tfrac{1}{4},\ \tfrac{1}{2} \text{ and } \tfrac{3}{4}, \end{aligned}$$

where N^* is the conventional single-stage sample size that is necessary to achieve the same power at $\lambda\xi_1$. The numerators of the ratios R_1 and R_2 are the expected sample size of a two-stage adaptive

Table 2.6: *Design parameters and their values of the response surface analysis for two-stage adaptive procedures.*

Design Parameter	Values
α_1	.01 .025 .04
β_1	.02 .05 .08
f	.3 .5 .75
ξ_1	.125 .2 .25
$A(y_1, \xi_0)$	$A_{Flat}(y_1, \xi_0) = a_{0F}$ $A_{LC}(y_1, \xi_0) = 1 - \Phi\left[\gamma\sqrt{n_1}\sqrt{(k_2 - \xi_0)^2 - (y_1 - \xi_0)^2}\right]$ $A_{Linear}(y_1, \xi_0) = a_0 + a_{1L}(y_1 - k_1)$ $A_{Step}(y_1, \xi_0) = a_0 + a_{1S}I(y_1 > (k_1 + k_2)/2)$ $A_{Root}(y_1, \xi_0) = a_0 + a_{1R}\sqrt{y_1 - k_1}$
$A(y_1, \xi_1)$	$A_{Flat}(y_1, \xi_1) = b_{0F}$ $A_{LC}(y_1, \xi_1) = 1 - \Phi\left[\gamma\sqrt{n_1}\sqrt{(\xi_1 - k_1)^2 - (\xi_1 - y_1)^2}\right]$ $A_{Linear}(y_1, \xi_1) = b_{0L} + b_{1L}(y_1 - k_1)$ $A_{Step}(y_1, \xi_1) = b_{0S} + b_{1S}I(y_1 > (k_1 + k_2)/2)$ $A_{Root}(y_1, \xi_1) = b_{0R} + b_{1R}\sqrt{y_1 - k_1}$

design. Therefore, R_1 compares a two-stage adaptive design to a conventional single-stage design that has the same power at the original ξ_1, and R_2 compares a two-stage adaptive design to a conventional single-stage design with the same power at the intermediate value, $\lambda\xi_1$. Power and expected sample size are numerically computed for each combination of design components.

To see which components cause relatively more variability in a response surface, an ANOVA type analysis is conducted. For each response, we analyze the ratios of the sum of squares of the main effects and the interaction effects to the total sum of squares. The model includes all six main effects and 15 two-way, 20 three-way, 15 four-way and 6 five-way interactions. Because this is essentially like a "one observation per cell" ANOVA design, the six-way interaction is not included in the model. We ignore the differences in the degrees of freedoms for the various sums of squares, and declare a particular effect to be "significant" if its sum of squares explain at least 5% of the total sum of squares.

Table 2.7 summarizes the results. Only significant effects (i.e., those with sums of squares accounting for more than 5% of the total sums of squares) are included in the table, with the exception of the effects due to $A(y_1, \xi_0)$ and $A(y_1, \xi_1)$. More detailed results are given in Appendix A. As Table 2.7 shows, the combination of f, α_1, β_1, $f \times \alpha_1$ and $f \times \beta_1$ explain at least 90% of the variation for each response variable: R_1, R_2 and the three powers.

A surprising result is that neither $A(y_1, \xi_0)$ nor $A(y_1, \xi_1)$ are "significant" factors. The biggest contribution that $A(y_1, \xi_0)$ makes is 4.2% of the total variation in R_1 at $\xi = \xi_1$. For the power, the

Table 2.7: *Proportions of sums of squares to the total sums of squares.*

Response	λ	f	α_1	β_1	$f \times \alpha_1$	$f \times \beta_1$	Total	$A(y_1, \xi_0)$	$A(y_1, \xi_1)$
R_1	0	.250	.189	.349	.094	.019	.901	.026	.022
	$\frac{1}{4}$	.371	.200	.220	.102	.020	.913	.018	.016
	$\frac{1}{2}$	.543	.131	.085	.111	.067	.937	.007	.006
	$\frac{3}{4}$	.570	.081	.107	.078	.099	.935	.019	.003
	1	.445	.144	.181	.028	.105	.903	.042	.009
	λ	f	α_1	β_1	$f \times \alpha_1$	$f \times \beta_1$	Total	$A(y_1, \xi_0)$	$A(y_1, \xi_1)$
R_2	$\frac{1}{4}$	.200	.418	.126	.197	.008	.949	.008	.000
	$\frac{1}{2}$	.253	.349	.138	.196	.002	.938	.009	.000
	$\frac{3}{4}$	.499	.164	.042	.181	.015	.901	.037	.005
	λ	f	α_1	β_1	$f \times \alpha_1$	$f \times \beta_1$	Total	$A(y_1, \xi_0)$	$A(y_1, \xi_1)$
Power	$\frac{1}{4}$	.057	.561	.130	.172	.037	.957	.006	.006
	$\frac{1}{2}$	.001	.425	.303	.124	.097	.950	.008	.007
	$\frac{3}{4}$	.045	.215	.470	.055	.162	.947	.009	.007

contribution of $A(y_1, \xi_0)$ is less than 1% everywhere. $A(y_1, \xi_1)$ is even less significant, its maximum contribution being only 2.2%. We have included various shapes for the $A(y_1, \xi_0)$ and $A(y_1, \xi_1)$ functions, and yet the effects of varying these shapes are much less significant than f, α_1 and β_1.

The findings above suggest that when we design a two-stage adaptive trial, we should carefully choose the design of Stage I, namely, the sample size, n_1, α_1=P(Reject in Stage I $\mid H_0$) and β_1=P(Accept in Stage I $\mid H_1$). The fact that the shapes of $A(y_1, \xi_0)$ and $A(y_1, \xi_1)$ are less important justifies altering these functions in designing process to meet specific design criteria such as controlling the maximum Stage II sample size.

In the example in Section 2.4.2, we start our planning of Stage II with $A(y_1, \xi_0)$ and $A(y_1, \xi_1)$. And we make modifications to the resulting $n_2(y_1)$ so that the $201 \leq n_2(y_1) \leq 327$. As a result, $A(y_1, \xi_1)$ is different from the originally planned one, but in light of the findings in this section, we conclude that this change has little effect on the design characteristics (power and expected sample size) of the design.

We now further summarize some general findings describing *how* these components affect the resulting design. Each "response surface" is complex and in addition there are 11 response variables (5 for R_1, 3 for R_2 and 3 for power). For each response variable, at a particular combination of f, α_1 and β_1, there are 75 combinations of A_0, A_1 and ξ_1. Because the variation due to A_0, A_1 and ξ_1 is found insignificant, we use the median of these 75 responses as the score for each of the 27 combinations of f, α_1 and β_1. For each of the 11 response variables, the rank of these 27 scores is used to assess the effect of f, α_1 and β_1. For R_1 and for R_2, a smaller value is more preferable, and the rank is given in ascending order; for the power, a larger value is more preferable, and the rank

is given in descending order. In other words, an ideal combination would have the rank of 1 for all 11 outputs.

Generally speaking, a small α_1 and a large β_1 seem to perform well. The best overall combination is $f = .5$, $\alpha_1 = .01$, $\beta_1 = .08$. This combination ranks 5th or better in 8 of 11 responses. The second best is $f = .75$, $\alpha_1 = .01$, $\beta_1 = .08$. This ranks 5th or better in 7 responses. The combination $f = .3$, $\alpha_1 = .01$, $\beta_1 = .08$ gives the best power at all three intermediate points, but the sample size for this combination, especially ξ near ξ_1, is among the worst.

It seems that as β_1 increases, the power increases, but so does the sample size. And as α_1 increases, the power and the sample size tend to decrease. However, the effect of α_1 is less clear than the effect of β_1. As seen in Table 2.7, the interaction effect between f and α_1 is fairy high. A small f generally gives very high power but the sample size is also large. It should also be noted that when f is large, there are very small differences in the performance characteristics for different choices of α_1 and β_1.

In conclusion, an intermediate to large f appears to be preferred to a small one. And a small to intermediate α_1 and an intermediate to large β_1 are preferred. Although $A(y_1, \xi_0)$ and $A(y_1, \xi_1)$ are found to have much smaller effect than α_1, β_1 and f, we did, for completeness, examine the effects of the A-functions in this study. One notable fact is that the combination of $A_{flat}(y_1, \xi_0)$ and $A_{flat}(y_1, \xi_1)$ performed much worse than any other combination of A-functions.

2.6 Discussion

In this chapter, we provided a framework to facilitate the design of two-stage adaptive procedures. This framework enables one to specify certain design components so as to construct a design which satisfies specific criteria with respect to unconditional Type I error rate, power and expected and maximum sample size.

For proper design of a two-stage adaptive procedure, the study protocol should clearly state all the actions to be taken at the end of Stage I. Some papers allow in an adaptive design for the clinical trial sponsor to choose the conditional power and the sample size of Stage II after observing y_1. From our perspective (and others, e.g., Liu et.al (2002)), in order to define the power function properly, these Stage II components need to be specified for all values of y_1 at the beginning of the trial.

The monotonicity of the power function of two-stage adaptive procedures appears to be often overlooked, but without this property the Type I errors rate need not be defined by the powers at the boundary of the null hypothesis. By specifying a nondecreasing $A(y_1, \xi_1)$ and nonincreasing $n_2(y_1)$,

we can guarantee that the power function increases monotonically up to ξ_1 and never goes below the power at ξ_1 when the alternative is greater than ξ_1.

Having this framework enables us to compare the effects of various design components. Our computation experiment shows that in constructing a two-stage adaptive design, Stage I specification components, α_1, β_1 and n_1 are relatively more important than the conditional power functions for Stage II despite the fact that the literature pays a fair amount of attention to the choice of conditional power functions. Our result suggests that the shape of the conditional power functions matters relatively little in determining design characteristics of a two-stage adaptive procedure.

Chapter 3

Practical Considerations For Two-Stage Adaptive Procedure

3.1 Introduction

In this chapter we present two important extensions of the framework for two-stage adaptive procedures developed in the previous chapter. In the construction of the framework, two restrictive assumptions have been made, namely equal sample sizes from both groups and known population variances. These assumptions served to make the development of the framework simple and straightforward. However, to be useful in practice, the framework should be able to handle those situations in which the assumptions of equal sample sizes and known variances cannot be made.

We first extend the framework to the case where the sample sizes from the two groups are not equal. It is sometimes advantageous to plan a two-stage adaptive design with unequal sample sizes from the control and treatment groups because of differences in the availability of observations from each group. Unequal sample sizes also result from missing observations. In Section 3.2.1, we provide a method for determining the sample sizes for a single-stage conventional study when the assumptions of balanced design and equal population variances cannot be made. Then, in Section 3.2.2, we show how to build a two-stage adaptive procedure with unbalanced sample sizes. Section 3.2.3 shows how to handle missing observations in Stage I, and Section 3.2.4 deals with missing observations in Stage II. Finally, Section 3.2.5 presents an example in which the techniques from the previous sections are utilized. We consider only those situations in which observations are missing completely at random. In treating missing observations, we still make the assumption of known population variances.

Another important practical extension is to unknown population variance, which is addressed in Section 3.3. In the construction of the original framework in Chapter 2, it was assumed that the two underlying populations were normally distributed with known variances, σ_t^2 for the treatment group and σ_c^2 for the control group. In this extension of the framework, we do assume that $\sigma_t^2 = \sigma_c^2 \equiv \sigma^2$, but we do not assume that the common variance, σ^2, is known. First we present a brief motivation

for using our framework. Several papers [e.g., Wittes and Brittain (1990) and Zucker, et.al. (1999)] suggest that we should break the blind to estimate the population variance, σ^2, without obtaining an estimate of the population means. In Section 3.3.1, we argue against this idea of "partially breaking the blind." In Section 3.3.2, we give a justification for not using the variance estimate from the combined-stage data. We consider only a variance estimate from a single stage, either Stage I or Stage II, in our framework. We show that if the sample size from one stage is moderately large, a variance estimate from this stage is almost as good as a variance estimate utilizing the data from both stages. The treatment of unknown variances is divided into two sections. First in Section 3.3.3, we suppose that the Stage I sample sizes, n_{1t} and n_{1c}, are large enough to treat s_{1p}^2, the pooled sample variance from Stage I observations, as an accurate estimate of σ^2. An example is presented in Section 3.2.5. Then in Section 3.3.5, we discuss designs with small Stage I sample sizes.

3.2 Unequal Sample Sizes

3.2.1 Sample Sizes For One-Stage Conventional Study

When determining the Stage I sample size for a two-stage adaptive design, we may use the one-stage conventional design's sample size as a guide. When the sample sizes from the treatment and control groups are allowed to be different and σ_t^2 and σ_c^2 are known but unequal, obtaining sample sizes that yield specific α and β is not straightforward even for the simple single-stage conventional study. In this section, we give a general form of the sample sizes for the single-stage design, N_t and N_c, and also derive the optimal sample sizes that yield the minimum total sample size.

As in Chapter 2, we suppose independent sampling from two normal populations with unknown means μ_t and μ_c and known variance σ_t^2 and σ_c^2, respectively, as below.

$$\begin{aligned} X_t &\sim \text{Normal}\left(\mu_t,\, \sigma_t^2\right), \\ X_c &\sim \text{Normal}\left(\mu_c,\, \sigma_c^2\right). \end{aligned} \tag{3.2.1}$$

Consider testing the following hypotheses:

$$\begin{aligned} H_0 &: \mu_t - \mu_c \leq 0 \\ H_1 &: \mu_t - \mu_c > 0. \end{aligned}$$

The sample sizes N_t and N_c are determined so that the Type I error rate is α and the Type II error rate is β at the alternative, $\mu_t - \mu_c = \Delta_1$. First we derive a general form for N_t and N_c, followed by a discussion of the optimal sample sizes. We have

$$\bar{X}_t - \bar{X}_c \sim \text{Normal}\left(\mu_t - \mu_c,\, \frac{\sigma_t^2}{N_t} + \frac{\sigma_c^2}{N_c}\right). \tag{3.2.2}$$

Let $\lambda = \dfrac{N_c}{N_t}$. We can then represent the variance in (3.2.2) as

$$\frac{\sigma_t^2}{N_t} + \frac{\sigma_c^2}{N_c} = \frac{2A^2}{M}, \tag{3.2.3}$$

where $A = \sqrt{(\lambda\sigma_t^2 + \sigma_c^2)/2}$ and $M = N_c$. We consider A as a "variance component" and M as a "sample size component." In order to have α and β as specified, we need

$$M = 2A^2 W, \tag{3.2.4}$$

where $W = (z_\alpha + z_\beta)^2/\Delta_1^2$. Using (3.2.3) and (3.2.4), we obtain

$$\frac{1}{W} = \frac{2A^2}{M} = \frac{\lambda\sigma_t^2}{N_c} + \frac{\sigma_c^2}{N_c} = \frac{1}{N_c}(\lambda\sigma_t^2 + \sigma_c^2),$$

which leads to

$$N_c = W\left(\lambda\sigma_t^2 + \sigma_c^2\right). \tag{3.2.5}$$

Substituting $\lambda = N_c/N_t$ and solving for N_c, we get

$$N_c = \frac{N_t\sigma_c^2 W}{N_t - \sigma_t^2 W}. \tag{3.2.6}$$

We can solve (3.2.6) for N_t to get

$$N_t = \frac{N_c\sigma_t^2 W}{N_c - \sigma_c^2 W}. \tag{3.2.7}$$

From (3.2.6) and (3.2.7), it is clear that N_c and N_t must be at least as large as $\sigma_c^2 W$ and $\sigma_t^2 W$, respectively. We can use any N_t and N_c as long as they satisfy (3.2.6) or equivalently (3.2.7).

Now, we discuss the minimum total sample size. Let $N_s = N_t + N_c$. We can write N_s as follows:

$$N_s = \left(\frac{1}{\lambda} + 1\right) N_c = \left(\frac{1}{\lambda} + 1\right) W\left(\lambda\sigma_t^2 + \sigma^2\right). \tag{3.2.8}$$

To obtain the minimum value of N_s, we differentiate (3.2.8) with respect to λ to get

$$\frac{\partial N_s}{\partial\lambda} = \left(-\frac{1}{\lambda^2}\right) W(\lambda\sigma_t^2 + \sigma_c^2) + \sigma_t^2 W\left(\frac{1}{\lambda} + 1\right). \tag{3.2.9}$$

Setting (3.2.9) equal to 0 and solving for λ, we get

$$\lambda_{opt} = \frac{\sigma_c}{\sigma_t}. \tag{3.2.10}$$

And $\partial^2 N_s/\partial\lambda^2$ evaluated at λ_{opt} is greater than 0, so (3.2.10) gives the minimum value of N_s.

Substituting (3.2.10) into (3.2.8) and (3.2.5) and because $N_t = N_c/\lambda$, we find that the optimal sample sizes, N_c^{opt} and N_t^{opt} that make the minimum total sample size, N_b^{opt}, are

$$N_c^{opt} = \sigma_c(\sigma_t + \sigma_c)W,$$
$$N_t^{opt} = \sigma_t(\sigma_t + \sigma_c)W.$$

And

$$N_s^{opt} = (\sigma_t + \sigma_c)^2 W.$$

3.2.2 Planned Unequal Sample Sizes

As in Chapter 2, we assume that the underlying distributions for the treatment and control groups are (3.2.1). And let n_{1t} and n_{1c} be the planned sample sizes of Stage I. Then the distributions of the sample means for Stage I are as follows :

$$\bar{X}_{1t} \sim \text{Normal}\left(\mu_t, \frac{\sigma_t^2}{n_{1t}}\right),$$
$$\bar{X}_{1c} \sim \text{Normal}\left(\mu_c, \frac{\sigma_c^2}{n_{1c}}\right).$$

Hence,

$$\bar{X}_{1t} - \bar{X}_{1c} \sim \text{Normal}\left(\mu_t - \mu_c, \frac{\sigma_t^2}{n_{1t}} + \frac{\sigma_c^2}{n_{1c}}\right), \tag{3.2.11}$$

Suppose that the ratio of the sample sizes from the treatment and control group, $\lambda_1 = n_{1c}/n_{1t}$ is given. As shown in Section 3.2.1, $\lambda_1 = \sigma_c/\sigma_t$ is a reasonable choice, but practically, λ_1 may be determined by outside factors such as availability and cost of the observations from each group. This ratio could be different in each stage, so we let λ_1 and λ_2 be the ratios of the sample sizes in Stage I and Stage II, respectively. Both λ_1 and λ_2 need to be determined prior to the start of Stage I.

In order to denote the variance of $\bar{X}_{1t} - \bar{X}_{1c}$ and $\bar{X}_{2t} - \bar{X}_{2c}$ in general, we define the following function:

$$V(\theta_1^2, \theta_2^2, n, \lambda) = \frac{\lambda\theta_1^2 + \theta_2^2}{n}.$$

Note that the variance of (3.2.11) is $V(\sigma_t^2, \sigma_c^2, n_{1c}, \lambda_1)$. In order to put the current formulation into the formulation of Chapter 2, we let:

$$V(\theta_1^2, \theta_2^2, n, \lambda) = \frac{2[a(\theta_1^2, \theta_2^2, \lambda)]^2}{n},$$

where $a(\theta_1^2,\, \theta_2^2,\, \lambda) = \sqrt{(\lambda\theta_1^2 + \theta_2^2)/2}$. To simplify the notation, let $a_1 = a(\sigma_t^2, \sigma_c^2, \lambda_1)$ and $m_1 = n_{1c}$. As in (2.2.1), we redefine Y_1 as follows:

$$Y_1 = \frac{\bar{X}_{1t} - \bar{X}_{1c}}{\sqrt{2}\, a_1}. \tag{3.2.12}$$

Then the distribution of Y_1 in (3.2.12) is

$$Y_1 \sim \text{Normal}\left(\xi_{\bullet 1},\, \frac{1}{m_1}\right), \tag{3.2.13}$$

where $\xi_{\bullet 1}$ is defined as $\Delta_\bullet/\sqrt{2}\, a_1$. The first subscript of ξ is associated with Δ and the second subscript is associated with a. For example, $\xi_{11} = \Delta_1/\sqrt{2}\, a_1$.

Conceptually, a_1^2 is a linear combination of σ_t^2 and σ_c^2, and m_1 serves as a new "sample size" for Stage I. Some special cases should be noted.

1. If $n_{1t} = n_{1c} \equiv n_1$, then let $a_1 = a(\sigma_t^2,\, \sigma_c^2,\, 1) = \sqrt{(\sigma_t^2 + \sigma_c^2)/2}$ and $m_1 = n_1$.
2. If $\sigma_t^2 = \sigma_c^2 \equiv \sigma^2$, then let $a_1 = a(\sigma^2,\, \sigma^2,\, \lambda_1) = \sqrt{\sigma^2(1 + \lambda_1)/2}$.
3. If $\sigma_t^2 = \sigma_c^2 \equiv \sigma^2$ and $n_{1t} = n_{1c} \equiv n_1$, then let $a_1 = a(\sigma^2,\, \sigma^2,\, 1) = \sigma$ and $m_1 = n_1$.

The original formulation of Chapter 2 is the special case 3 above. For planning Stage I, we can use the method of Chapter 2 with σ^2 and n_1 replaced by a_1^2 and m_1, respectively. The actual sample sizes are $n_{1t} = m_1/\lambda_1$ and $n_{1c} = m_1$.

For Stage II, let

$$m_2(y_1) = n_{2c}(y_1).$$

And define Y_2, similarly to (2.2.3), as follows:

$$Y_2 = \frac{\bar{X}_{2t} - \bar{X}_{2c}}{\sqrt{2}\, a_2}, \tag{3.2.14}$$

where $a_2 = a(\sigma_t^2,\, \sigma_c^2,\, \lambda_2)$. The distribution of Y_2 given that $Y_1 = y_1$ is Normal $(\xi_{\bullet 2},\, 1/m_2(y_1))$ where $\xi_{\bullet 2}$ is defined as $\Delta_\bullet/\sqrt{2}\, a_2$.

The method of Chapter 2 is again applicable with $n_2(y_1)$ replaced by $m_2(y_1)$. The actual sample sizes from both groups are then calculated as $n_{2t}(y_1) = m_2(y_1)/\lambda_2$ and $n_{2c} = m_2(y_1)$. There is a small additional complexity when $\lambda_1 \neq \lambda_2$, (unless $\Delta_\bullet = 0$) because then $\xi_{\bullet 1} \neq \xi_{\bullet 2}$. When specifying Stage I components, we use $\xi_{\bullet 1}$. For example, $\alpha_1 = 1 - \Phi\left[\sqrt{m_1}\,(k_2 - \xi_{01})\right]$. And for Stage II components, we use $\xi_{\bullet 2}$. For example, $A(y_1, \xi_{02}) = 1 - \Phi\left[\sqrt{m_2(y_1)}\,(w(y_1) - \xi_{02}\right]$. When $\lambda \equiv \lambda_1 = \lambda_2$, the second subscript on the ξ's introduced in this section can be omitted. Letting $\lambda \equiv \lambda_1 = \lambda_2$ simplifies notation and computation without losing much generality, therefore, in the following sections, we assume that $\lambda \equiv \lambda_1 = \lambda_2$.

3.2.3 Missing Observations In Stage I

In this section and the next we consider situations where observations are missing completely at random. We assume that the population variances, σ_t^2 and σ_c^2, are known.

Suppose that sample sizes of n_{1t} and n_{1c} from the treatment and control groups are planned for Stage I and that some observations are missing completely at random. Let n'_{1t} and n'_{1c} be the actual number of observations. While n'_{1t} and n'_{1c} are random variables, their distributions are assumed not to depend on any parameters. Hence, we can treat them as constants (conditioned on their observed values). Let $\lambda = n_{1c}/n_{1t}$. With the change of sample sizes, λ also changes. Let $\lambda'_1 = n'_{1c}/n'_{1t}$. Let Y'_1 be defined like Y_1 in (3.2.12):

$$Y'_1 = \frac{\bar{X}_{1t} - \bar{X}_{1c}}{\sqrt{2}\, a}, \tag{3.2.15}$$

where $a = a(\sigma_t^2,\, \sigma_c^2,\, \lambda) = \sqrt{(\lambda\sigma_t^2 + \sigma_c^2)/2}$. The distribution of Y'_1 is as follows:

$$Y'_1 \sim \text{Normal}\left(\xi,\, \frac{1}{m'_1}\right). \tag{3.2.16}$$

The new "sample size," m'_1, satisfies the following:

$$\frac{2a^2}{m'_1} = \frac{n'_{1c}\sigma_t^2 + n'_{1t}\sigma_c^2}{n'_{1t}n'_{1c}} \tag{3.2.17}$$

Solving (3.2.17) for m'_1, we get

$$\begin{aligned} m'_1 &= \frac{n'_{1t}n'_{1c}(\lambda\sigma_t^2 + \sigma_c^2)}{n'_{1c}\sigma_t^2 + n'_{1t}\sigma_c^2} \\ &= n'_{1c}\frac{\lambda\sigma_t^2 + \sigma_c^2}{\lambda'_1\sigma_t^2 + \sigma_c^2}. \end{aligned} \tag{3.2.18}$$

Because m'_1 arises due to missing observations, it follows that $m_1 > m'_1$. This can be verified as follows:

$$\begin{aligned} m'_1 &= n'_{1c}\frac{(n_{1c}/n_{1t})\,\sigma_t^2 + \sigma_c^2}{(n'_{1c}/n'_{1t})\,\sigma_t^2 + \sigma_c^2} \\ &= n_{1c}\,\frac{\sigma_t^2/n_{1t} + \sigma_c^2/n_{1c}}{\sigma_t^2/n'_{1t} + \sigma_c^2/n'_{1c}} \\ &\le n_{1c} = m_1 \end{aligned}$$

The inequality is true because $n'_{1t} \le n_{1t}$ and $n'_{1c} \le n_{1c}$ with at least one strict inequality holding. The equality, $m'_1 = m_1$, only holds when there are no missing observations.

Obviously, the distribution of Y'_1 is different from the distribution of the original Y_1. The relationship between Y_1 and Y'_1 is,

$$Y'_1 = \frac{\sqrt{m_1}}{\sqrt{m'_1}}\,(Y_1 - \xi) + \xi. \tag{3.2.19}$$

Before Stage I, k_1, k_2 are calculated so that α_1 and β_1 are the Stage I error probabilities based on the distribution of Y_1. When there are missing observations we handle this by modifying k_1 and k_2 to reflect the distribution of Y_1'. We note that modifying k_1 and k_2 is not the only way to respond to the change of sample size in Stage I. For example, we could keep the same k_1 and k_2 and calculate new α_1 and β_1, and change $A(y_1, \xi_0)$ to keep the overall α at the targeted level. However, we feel that to change the critical values, k_1 and k_2, rather than α_1 and β_1, is a simpler way to deal with missing observations. Also a potential problem with keeping the original k_1 and k_2 is that the actual Stage I error probabilities may exceed the targeted overall error probabilities.

Recall that k_1 is determined in order to set β_1, which depends on the distribution under the original alternative, and k_2 determines α_1, which depends on the null distribution. Thus ξ in (3.2.19) is ξ_1 for the transformation of k_1, and ξ_0 for the transformation of k_2:

$$\begin{aligned} k_1' &= \frac{\sqrt{m_1}}{\sqrt{m_1'}}(k_1 - \xi_1) + \xi_1, \\ k_2' &= \frac{\sqrt{m_1}}{\sqrt{m_1'}}(k_2 - \xi_0) + \xi_0. \end{aligned} \tag{3.2.20}$$

By these transformations, the Stage I error probabilities, α_1 and β_1, are maintained at the targeted values.

When k_1 and k_2 are modified, the Stage II components need to be modified to keep the Type I error rate at α and the power at $1 - \beta$. To preserve the integrity of the adaptive procedure, one needs to describe in the protocol how to modify the design of Stage II at the end of Stage I. Let $A'(y_1', \xi)$ be the new conditional power function at ξ, and let $m_2'(y_1')$ be the new Stage II sample size function and let $w'(y_2')$ be the new critical value function.

We first establish the following fact with regards to the ordering of k_1 and k_1' and k_2 and k_2'. We assume that $k_1 \neq -\infty$ and $k_2 \neq \infty$.

Fact 3.2.1. *When there are missing observations in Stage I, $k_1' < k_1$ and $k_2 < k_2'$.*

The proof is straightforward from the facts that follow from (3.2.20), that is, $\sqrt{m_1'}\,(k_1' - \xi_1) = \sqrt{m_1}\,(k_1 - \xi_1)$ and $k_1 < \xi_1$, $k_1' < \xi_1$; $\sqrt{m_1'}\,(k_2' - \xi_0) = \sqrt{m_1}\,(k_2 - \xi_0)$ and $k_2 > \xi_0$, $k_2' > \xi_0$. Thus, the continuation region is extended when $m_1' < m_1$.

There are numerous ways to modify the Stage II specification components, but we present one straightforward approach. Let $A'(y_1', \xi_0)$ have the following forms:

$$A'(y_1', \xi_0) = \begin{cases} \theta_0 A(k_1, \xi_0) & \text{if } k_1' < y_1' \leq k_1, \\ \theta_0 A(y_1', \xi_0) & \text{if } k_1 < y_1' \leq k_2, \\ \theta_0 A(k_2, \xi_0) & \text{if } k_2 < y_1' \leq k_2'. \end{cases} \tag{3.2.21}$$

$A'(y_1', \xi_0)$ is an extension of $A(y_1, \xi_0)$ from the domain $(k_1,\, k_2\,]$ to the domain $(k_1',\, k_2'\,]$ through rescaling (by θ_0) over $(k_1,k_2]$, and a constant extension beyond. The value θ_0 is chosen to adjust $A(y_1, \xi_0)$ so that the Type I error rate is α. In addition, let $m_2'(y_1')$ have the following form:

$$m_2'(y_1') = \begin{cases} m_2(k_1) + \eta & \text{if } k_1' < y_1' \le k_1, \\ m_2(y_1') + \eta & \text{if } k_1 < y_1' \le k_2, \\ m_2(k_2) + \eta & \text{if } k_2 < y_1' \le k_2', \end{cases} \tag{3.2.22}$$

where η is a non-negative number. The new "sample size," $m_2'(y_1')$, is a constant extension to the broader domain $(k_1',\, k_2'\,]$ and an upward shift by some constant, η. Given (3.2.21) and (3.2.22), we can obtain $A'(y_1', \xi_1)$ and $w'(y_2')$ through the relationships that these four components must satisfy. We can adjust η in (3.2.22) so that the resulting $A'(y_1', \xi_1)$ satisfies the condition that the power is $1 - \beta$ at Δ_1:

$$\rho_2 = \int_{k_1'}^{k_2'} A'(y_1', \xi_1) g'_{\xi_1}(y_1')\, dy_1', \tag{3.2.23}$$

where $g'_{\xi_1}(y_1')$ is the pdf of Normal $(\xi_1,\, 1/m_1')$. With this approach, we can also control the maximum Stage II sample size by introducing an upper bound for (3.2.22). It is possible that the power $1 - \beta$ is not attainable with a limit on the Stage II sample size. If so, we then take the maximum number of observations allowed, regardless of y_1'.

In this particular method, we choose $A'(y_1', \xi_0)$ and $m_2'(y_1')$ to be the two specifying components of Stage II. One could develop similar procedures for specifying any two components among $A'(y_1', \xi_0)$, $A'(y_1', \xi_1)$, $m_2'(y_1')$ and $w'(y_2')$. However, we feel that $A'(y_1', \xi_0)$ and $m_2'(y_1')$ are important because they give the sponsor direct control of the Type I error rate and the sample size. As in Stage I, the actual sample sizes are obtained from $m_2'(y_1')$ as $n_{2t}'(y_1') = m_2'(y_1')/\lambda_2$ and $n_{2c}'(y_1') = m_2'(y_1')$. With the approach described above, the form of all the elements of Stage II can be completely determined for any y_1'.

3.2.4 Missing Observations In Stage II

Now suppose further that some observations in Stage II are missing completely at random. We illustrate the Stage II procedure assuming that there are no missing observations in Stage I. However, if there are some missing observations in Stage I, by applying the method described in Section 3.2.3 we are essentially in the situation of no missing observations in Stage I. Suppose that we observe, in Stage I, $Y_1 = t$, where $k_1 < t \le k_2$. Then we plan to use $\nu_2 \equiv m_2(t)$ for the Stage II "sample size." For each group, the planned sample sizes are $\nu_{2t} = \nu_2/\lambda_2$ and $\nu_{2c} = \nu_2$. Let ν_{2t}' and ν_{2c}' be the actual sample sizes of Stage II, and let $\lambda_2' = \nu_{2c}'/\nu_{2t}'$. Let Y_2' be defined like Y_2 in (3.2.14) as follows:

$$Y_2' = \frac{\bar{X}_{2t} - \bar{X}_{2c}}{\sqrt{2}\, a},$$

where $a = a(\sigma_t^2,\, \sigma_c^2,\, \lambda) = \sqrt{(\lambda\sigma_t^2 + \sigma_c^2)/2}$. Given that $Y_1 = t$ and given the actual sample sizes of ν'_{2t} and ν'_{2c}, the distribution of Y'_2 is:

$$Y'_2 \sim \text{Normal}\left(\xi,\, \frac{1}{\nu'_2}\right),$$

where the new "sample size," ν'_2, satisfies the following:

$$\frac{2a^2}{\nu'_2} = \frac{\nu'_{2c}\sigma_t^2 + \nu'_{2t}\sigma_c^2}{\nu'_{2t}\nu'_{2c}}\,. \tag{3.2.24}$$

Solving (3.2.24) for ν'_2, we get

$$\nu'_2 = \nu'_{2c}\frac{\lambda\sigma_t^2 + \sigma_c^2}{\lambda'_2\sigma_t^2 + \sigma_c^2}\,. \tag{3.2.25}$$

An argument similar to that used in Section 3.2.3 to show that $m'_1 < m_1$, it can be shown that $\nu'_2 < \nu_2$ when there are missing observations in Stage II.

The Type I error rate of the procedure is protected simply by keeping $A(y_1, \xi_0)$ unchanged because the power function evaluated at ξ_0 does not depend on the Stage II sample size. The sample size of Stage II cannot be changed because a sample had already been taken. By Fact 2.2.2, changing the "sample size" from the planned ν_2 to the actual ν'_2 while keeping $A(y_1, \xi_0)$ unchanged will alter $A(t, \xi^*)$ for all $\xi^* \neq \xi_0$. More precisely, by (2.2.17), we can write

$$z_{A(t,\,\xi_0)} = z_{A(t,\,\xi^*)} + \sqrt{\nu_2}\,(\xi^* - \xi_0).$$

Therefore,

$$\begin{aligned} A'(t,\,\xi^*) &= 1 - \Phi\left[z_{A(t,\,\xi_0)} - \sqrt{\nu'_2}\,(\xi^* - \xi_0)\right] \\ &= 1 - \Phi\left[z_{A(t,\,\xi^*)} + \sqrt{\nu_2}\,(\xi^* - \xi_0) - \sqrt{\nu'_2}\,(\xi^* - \xi_0)\right] \\ &= 1 - \Phi\left[z_{A(t,\,\xi^*)} + (\sqrt{\nu_2} - \sqrt{\nu'_2}\,)(\xi^* - \xi_0)\right]. \end{aligned} \tag{3.2.26}$$

Similarly, because $z_{A(t,\,\xi_0)} = \sqrt{\nu_2}\,(w(t) - \xi_0)$,

$$\begin{aligned} w'(t) &= \frac{z_{A(t,\,\xi_0)}}{\sqrt{\nu'_2}} + \xi_0 \\ &= \frac{\sqrt{\nu_2}\,(w(t) - \xi_0)}{\sqrt{\nu'_2}} + \xi_0 \\ &= \sqrt{\frac{\nu_2}{\nu'_2}}\; w(t) + \left(1 - \sqrt{\frac{\nu_2}{\nu'_2}}\right)\xi_0. \end{aligned} \tag{3.2.27}$$

From (3.2.26), when there are missing observations in Stage II and hence $\nu'_2 < \nu_2$, $A'(t,\, \xi^*) < A(t,\, \xi^*)$ for all $\xi^* > \xi_0$. In other words, the conditional power at $\xi^* > \xi_0$ is smaller than what was originally

planned. The conditional Type I error rate, $A'(t, \xi_0)$ is equal to $A(t, \xi_0)$, and for $\xi < \xi_0$, the conditional power increases when $\nu_2' < \nu_2$, but even so, for all $\xi < \xi_0$, we have $A'(t, \xi) < A(t, \xi_0)$. Therefore, the Type I error rate for a composite null hypothesis is protected at α.

While we do not endorse the concept of post-hoc power, there may be circumstances when it needs to be computed for an adaptive test. In order to calculate honest post-hoc power, one theoretically needs to know what the "missingness" would have been for y_1 values other than the one observed. The reason, as noted in Chapter 2, is that in order to calculate the overall unconditional power, one needs to specify the Stage II sample size for all values of $y_1 \in (k_1, k_2]$. We present two different reasonable approaches to obtain the Stage II sample size as a *function* of $y_1 \in (k_1, k_2]$ when there is missingness in Stage II.

First, we could assume that the number of missing observations would be proportional to the intended sample size regardless of the value of y_1. We do not consider the proportion of missing observations within each of the treatment and control groups. Instead of using v_{2t}'/v_{2t} or v_{2c}'/v_{2c}, we use ν_2'/ν_2 as the ratio of number of non-missing observations to intended number of observations. With this assumption, we can write the new Stage II sample size function as $m_2^*(y_1) = \dfrac{\nu_2'}{\nu_2}\, m_2(y_1)$. Alternatively, the Stage II sample size may be truncated to ν_2'. In this case we assume that the new sample size function, $m_2^*(y_1)$, is,

$$m_2^*(y_1) = \begin{cases} m_2(y_1) & \text{if } m_2(y_1) < \nu_2', \\ \nu_2' & \text{otherwise.} \end{cases}$$

No matter which of the two approaches is used to obtain $m_2^*(y_1)$, we can write, by Fact 2.2.2, the Stage II critical value, $w'(y_1)$, and the conditional power function, $A'(y_1, \xi)$ at any ξ, as functions of $m_2^*(y_1)$ and $A(y_1, \xi_0)$ as follows.

$$\begin{aligned} w'(y_1) &= \frac{z_{A(y_1,\xi_0)}}{\sqrt{m_2^*(y_1)}} + \xi_0, \\ A'(y_1, \xi) &= 1 - \Phi\left[z_{A(y_1,\xi_0)} - \sqrt{m_2^*(y_1)}\,(\xi - \xi_0) \right]. \end{aligned} \tag{3.2.28}$$

Because the conditional power at $\xi \neq \xi_0$ is different from what was planned, the unconditional power at the original alternative, ξ_1, is no longer $1 - \beta$. We can calculate the post-hoc power of this procedure at any fixed ξ by

$$1 - \Phi\left[\sqrt{m_1}\,(k_2 - \xi)\right] + \int_{k_1}^{k_2} A'(y_1, \xi) g_\xi(y_1)\, dy_1. \tag{3.2.29}$$

We note that this post-hoc power approach is very dependent on the assumption one makes about the unobserved values of y_1. For instance, if one were to make the unrealistic assumption that for all values other than the observed y_1, the sample size did not change, then the post-hoc power would be the same as the planned power because $A(y_1, \xi_1) = A'(y_1, \xi_1)$ almost everywhere.

3.2.5 Example

In this section, we present an example that illustrates the techniques discussed in Sections 3.2.2, 3.2.3 and 3.2.4. Suppose that the hypotheses of interest are $H_0 : \mu_t - \mu_c \leq 0$ and $H_1 : \mu_t - \mu_c > 0$. The error probabilities are set to be $\alpha = .025$, and $\beta = .10$ at $\mu_t - \mu_c = \Delta_1 = 1$. And suppose that the known population variances are $\sigma_t^2 = 5^2$ and $\sigma_c^2 = 3^2$. Further suppose that the ideal ratio of control to treatment sample sizes is $1 : 2$, i.e., $\lambda = 1/2$. Given this λ, we can calculate $a = \sqrt{(25/2 + 9)/2} = 3.279$ and $\xi_1 = 1/(\sqrt{2} \times 3.279) = .2157$. Note that $\xi_0 = 0$.

First, we find the optimal sample sizes for the one-stage conventional procedure. As in Section 3.2.1, let $W = (z_\alpha + z_\beta)^2/\Delta_1^2$. For this example, we have $W = (z_{.025} + z_{.10})^2/1^2 = (1.96 + 1.28)^2/1 = 10.5$. Thus, by (3.2.7), we can use any N_t and N_c as long as they satisfy

$$N_t = \frac{N_c(25)(10.5)}{N_c - (9)(10.5)} = \frac{262.5\, N_c}{N_c - 94.5}.$$

In order to minimize the total sample size, the ratio of sample sizes, N_c/N_t needs to be $\lambda_{opt} = \sigma_c/\sigma_t = 3/5$. The sample sizes that make the total sample size minimum are

$$N_t^{opt} = \sigma_t(\sigma_t + \sigma_c)W = 5(5 + 3)(10.5) = 420,$$
$$N_c^{opt} = \sigma_c(\sigma_t + \sigma_c)W = 3(5 + 3)(10.5) = 252.$$

A single-stage conventional study's sample sizes are plotted as a function of λ in Figure 3.1. As shown before, the minimum total sample size occurs at $\lambda_{opt} = 3/5$, and it is $420 + 252 = 672$. As the plot shows, the necessary total sample size decreases sharply for λ in $(.1, .4)$ and it is fairly flat for λ in $(.4, .8)$. Therefore, even though the given restriction, $\lambda = .5$ does not give the most effective sampling plan, it is very close to the ideal situation, $\lambda = .6$.

For $\lambda = .5$, we can calculate the one-stage conventional study's sample sizes using (3.2.5), and they are $N_c = 10.5(.5 \times 5^2 + 3^2) \approx 226$ and $N_t = 226/.5 = 452$. Let the maximum sample sizes allowed be a 30% increase from N_t and N_c. The restrictions are $N_t^{max} = N_t \times 1.3 = 588$ and $N_c^{max} = N_c \times 1.3 = 294$.

Suppose that 75% of N_t and N_c are taken in Stage I. So we have $n_{1t} = 340$ and $n_{1c} = 170$. Further suppose that the Stage I specification components, in addition to $m_1 = n_{1c} = 170$ are $\alpha_1 = .005$ and $\beta_1 = .05$. With these three specification components, the remaining two can be calculated.

$$k_1 = \xi_1 - \frac{z_{\beta_1}}{\sqrt{m_1}} = .2157 - 1.645/\sqrt{170} = .0895,$$
$$k_2 = \xi_0 + \frac{z_{\alpha_1}}{\sqrt{m_1}} = 0 + 2.576/\sqrt{170} = .1976.$$

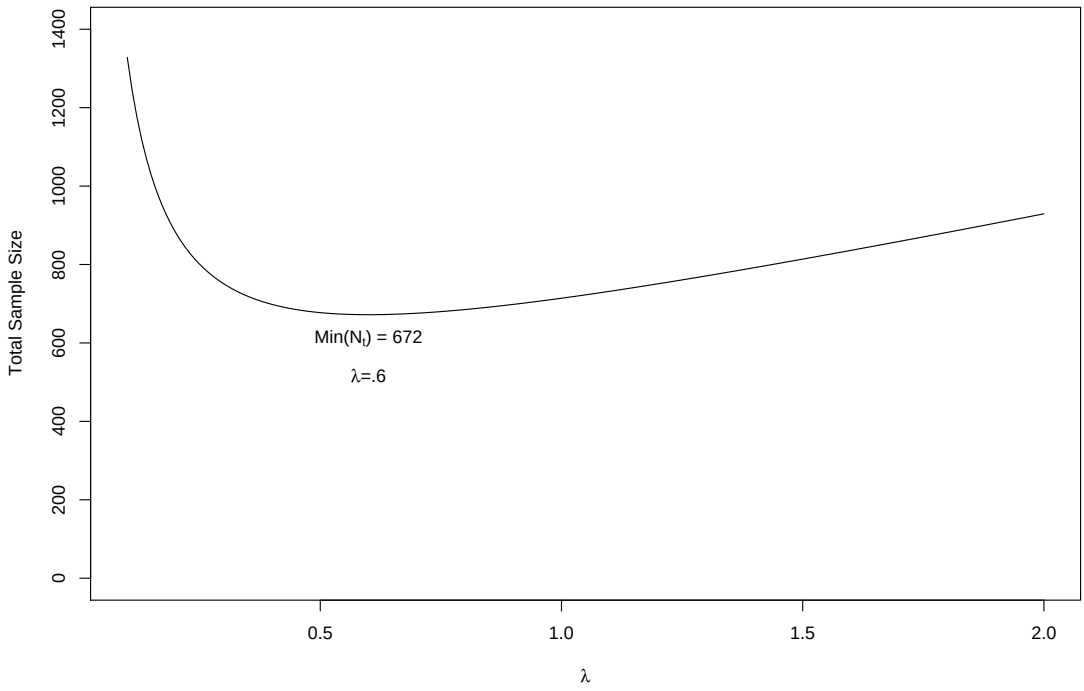


Figure 3.1: *Total sample size for one-stage conventional procedure as a function of* λ.

Thus, the five specification components for this particular Stage I are $\{\, \alpha_1 = .005,\, \beta_1 = .05,\, m_1 = 170,\, k_1 = .0895,\, k_2 = .1976 \,\}$.

For Stage II specification components, we might choose the following A functions.

$$A(y_1, \xi_0) = .08 + 49.8(y_1 - k_1)^2, \tag{3.2.30}$$
$$A(y_1, \xi_1) = .80 + 11.8(y_1 - k_1)^2.$$

These two functions give the correct Type I error rate and the power at Δ_1, and they are shown in Figure 3.2. The resulting sample size function is also displayed in Figure 3.3. It is decreasing in y_1 in the entire range of y_1 in $(k_1, k_2]$ and thus satisfies the condition of Lemma 2.3.2. The maximum Stage II sample sizes are $n_{2t}(k_1) = 218$ and $n_{2c}(k_1) = 109$ resulting in the maximum total sample sizes of $340 + 218 = 558$ and $170 + 109 = 279$ for the treatment and control groups, respectively. These values are below the maximum allowed sample sizes of 588 and 294. Table 3.1 shows Stage I probabilities, power and expected sample sizes for various values of Δ.

Now we specify how to handle missing observations in Stage I and in Stage II. If there are any missing observations in Stage I, the critical values, k_1 and k_2, are transformed using (3.2.20) to k_1' and k_2' in order to protect α_1 and β_1. With this change of Stage I critical values, Stage II specification components need to be modified. We use the transformation of $A(y_1, \xi_0)$ as described in (3.2.21)

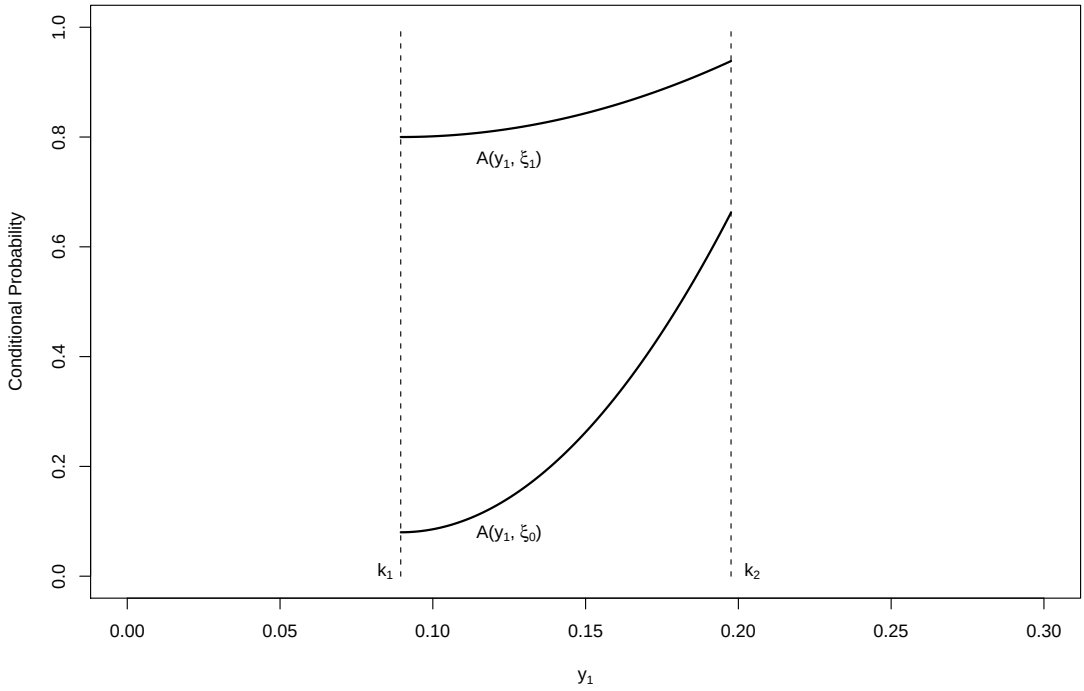


Figure 3.2: *Example of Section 3.2.5: $A(y_1, \xi_0)$ and $A(y_1, \xi_1)$ for the original design.*

and the transformation of $m_2(y_1)$ in (3.2.22) to obtain $A'(y_1', \xi_0)$ and $m_2'(y_1')$. Missing observations in Stage II are handled by calculating new critical value and conditional power using formulas in (3.2.27) and (3.2.26).

Suppose that by the end of Stage I, 44 observations are missing at random from the treatment group, 29 observations are missing at random from the control group. We have $n_{1t}' = 296$ and $n_{1c}' = 141$. In order to keep α_1 and β_1 as planned, we modify k_1 and k_2. First we calculate $\lambda_1' = 141/296 = .4764$. From (3.2.18), we get

$$m_1' = 141 \frac{.5 \times 5^2 + 3^2}{.4764 \times 5^2 + 3^2} = 145.0.$$

And from (3.2.20), the modified Stage I critical values are

$$k_1' = \frac{\sqrt{170}}{\sqrt{145.0}} (.0895 - .2157) + .2157 = .0791,$$
$$k_2' = \frac{\sqrt{170}}{\sqrt{145.0}} (.1976) = .2140.$$

In order to modify $A(y_1, \xi_0)$ to $A'(y_1', \xi_0)$ we need to find the appropriate θ_0 in (3.2.21). In order to do that, we temporary let $\theta_0 = 1$, and we find that P[Reject H_0 in Stage II] = .029. Because we need this probability to be $\alpha_2 = .025 - .005 = .020$, the correct θ is .020/.029 = .6897. And from

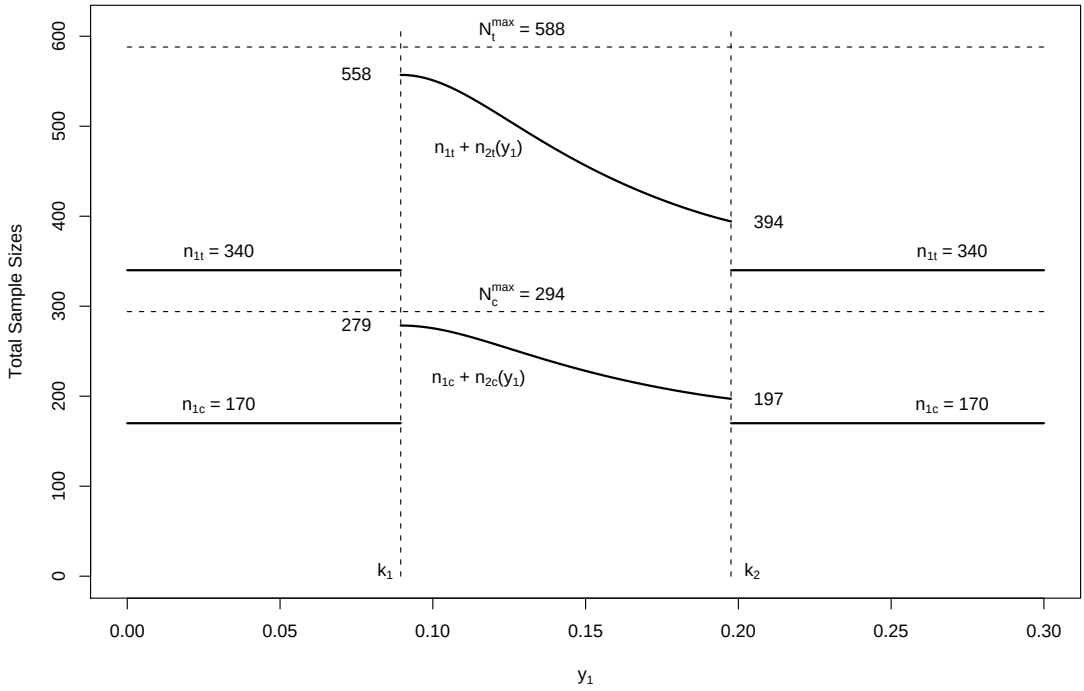


Figure 3.3: *Example of Section 3.2.5: Total Sample Sizes for the original design.*

(3.2.30), $A(k_1, \xi_0) = .08$ and $A(k_2, \xi_0) = .6624$, and we have the following result.

$$A'(y_1', \xi_0) = \begin{cases} .6897 \times .08 & \text{if } k_1' < y_1' \le k_1, \\ .6897\, A(y_1', \xi_0) & \text{if } k_1 < y_1' \le k_2, \\ .6897 \times .6624 & \text{if } k_2 < y_1' \le k_2'. \end{cases}$$

Stage II "sample size," $m_2'(y_1')$, is obtained as follows. First we use (3.2.22) with $\eta = 0$ to obtain the general form of $m_2'(y_1')$. This $m_2(y_1)$ is extended by flat lines for y_1' in (k_1', k_1) and in (k_2, k_2'). Then η determines how much upward shift is added to this basic shape. Also note that the maximum allowed for $m_2'(y_1')$ is $294 - 170 = 124$. With various η, we have different $m_2'(y_1')$ function, and we can obtain $A'(y_1', \xi_1)$ using the connection among the four specification components. We find that $A'(y_1', \xi_1)$ resulting from $m_2'(y_1')$ with the choice $\eta = 33$ satisfies the unconditional power condition,

$$\psi(\xi_1) = \int_{-\infty}^{\infty} A'(y_1', \xi_1) g_{\xi_1}(y_1')\, dy_1' = .90,$$

where $g_{\xi_1}(y_1')$ is the pdf of Normal $(\xi_1,\, 1/m_1')$. Thus, we use $\eta = 33$. And we also compute, for this particular η, to find that $m_2'(y_1') = 124$ (maximum allowed) for all y_1' that is smaller than .1170. Thus,

$$m_2'(y_1') = \begin{cases} 124 & \text{if } k_1' < y_1' \le .1170, \\ m_2(y_1') & \text{if } .1170 < y_1' \le k_2, \\ 60 & \text{if } k_2 < y_1' \le k_2'. \end{cases} \tag{3.2.31}$$

Table 3.1: *Example of Section 3.2.5: Stage I probabilities, power and expected sample sizes with no missing observations.*

Δ	Stage I			Stage II	Power	Sample Size	
	Accept	Continue	Reject	Reject		Control	Treatment
0.00	.878	.117	.005	.020	.025	179.8	359.6
0.25	.678	.291	.031	.098	.130	192.2	385.4
0.50	.405	.474	.121	.255	.376	203.7	407.4
0.75	.173	.507	.320	.369	.689	202.5	405.0
1.00	.050	.357	.593	.307	.900	190.5	381.0
1.25	.009	.165	.826	.154	.980	178.5	357.0

The last two columns are expected total sample sizes of 2 stages for the control and treatment groups.

The number 60 is $\eta + m_2(k_2)$. The actual sample sizes are $n_{2t}(y_1') = \lambda m_2'(y_1')$ and $n_{2c}(y_1') = m_2'(y_1')$. See Figure 3.4 for the new A functions and Figure 3.5 for the new sample sizes.

Table 3.2: *Example of Section 3.2.5: Stage I probabilities, power and expected sample sizes after accommodating Stage I missing observations.*

Δ	Stage I			Stage II	Power	Sample Size	
	Accept	Continue	Reject	Reject		Control	Treatment
0.00	.830	.165	.005	.020	.025	184.1	368.2
0.25	.619	.354	.027	.101	.128	197.8	395.6
0.50	.364	.535	.101	.275	.376	207.9	415.8
0.75	.159	.576	.265	.424	.689	206.0	412.0
1.00	.050	.442	.508	.392	.900	194.1	388.2
1.25	.011	.241	.748	.231	.979	181.5	363.0

The last two columns are expected total sample sizes of 2 stages for the control and treatment groups.

Now suppose that we observe, in Stage I, that $y_1' = .15$. Because $k_1' < y_1' < k_2'$, we move on to Stage II. From (3.2.31), we find that $\nu_2 = m_2'(.15) = 91$. So the actual sample sizes for Stage II are $\nu_{2t} = n_{2t}(.15) = 182$ and $\nu_{2c} = n_{2c}(.15) = 91$. Moreover, our calculation shows that the Stage II critical value is $w(.15) = .0954$.

Then suppose that in Stage II, the numbers of missing observations are 33 from the treatment group and 18 from the control group, respectively. Then $\nu_{2t}' = 149$ and $\nu_{2c}' = 73$, which leads to $\lambda_2' = 73/149 = .4899$. Using (3.2.25), we get

$$\nu_2' = 73\frac{.5 \times 5^2 + 3^2}{.4899 \times 5^2 + 3^2} = 73.9.$$

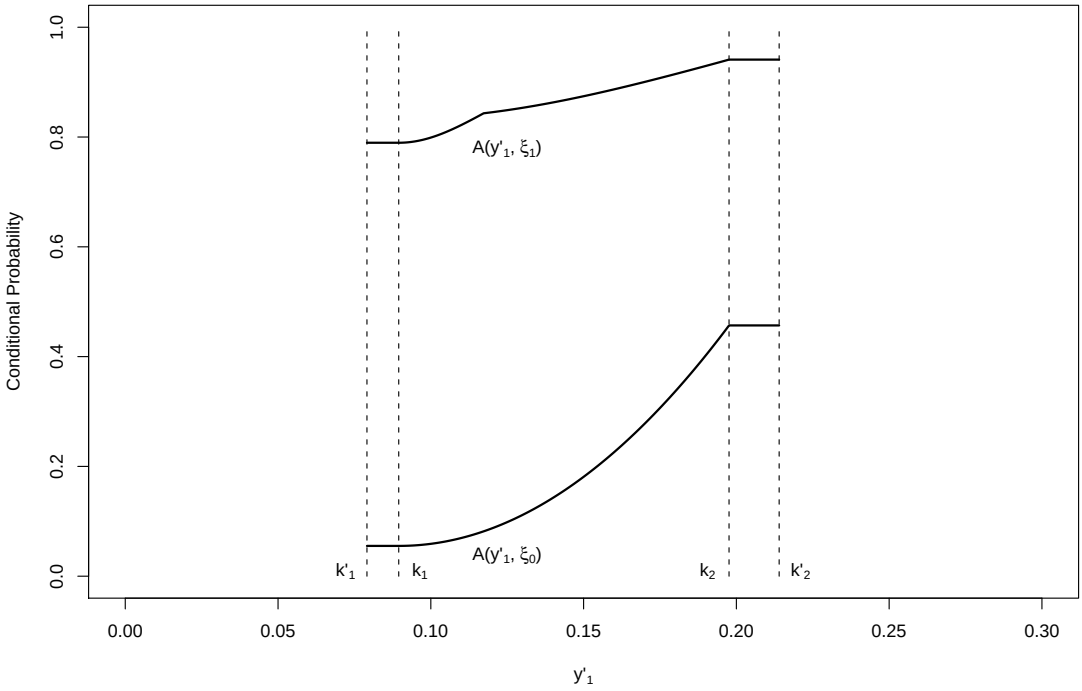


Figure 3.4: *Example of Section 3.2.5: $A(y_1, \xi_0)$ and $A(y_1, \xi_1)$ after accommodating missing observations in Stage I.*

Using (3.2.26), we can calculate the conditional power at any ξ. At the original alternative, ξ_1,

$$\begin{aligned} A'(.15,\ .2157) &= 1 - \Phi\left[z_{A(.15,\ .2157)} + (\sqrt{91}\ - \sqrt{73.9}\,)(.2157 - 0)\right] \\ &= 1 - \Phi\left[z_{.8650} + .9429 \times .2157\right] \\ &= 1 - \Phi\left[-1.1029 + .2034\right] = 1 - \Phi\left[-.8995\right] = .8158. \end{aligned}$$

The conditional Type I error rate does not change from

$$A(.15,\ 0) = .1738.$$

Moreover, using (3.2.27), we can calculate the new critical value as follows:

$$\begin{aligned} w'(.15) &= \sqrt{\frac{91}{73.9}}\ w(.15) + \left(1 - \sqrt{\frac{91}{76.7}}\ \right) \times 0 \\ &= \sqrt{1.231}\ \times 0.0954 = .1059. \end{aligned}$$

So we would reject H_0 in Stage II if $y_2' > .1059$.

We now calculate the post-hoc power at the original alternative, ξ_1, using the two assumptions on missingness discussed in Section 3.2.4. The first of the assumptions is that ν_2'/ν_2 stays constant

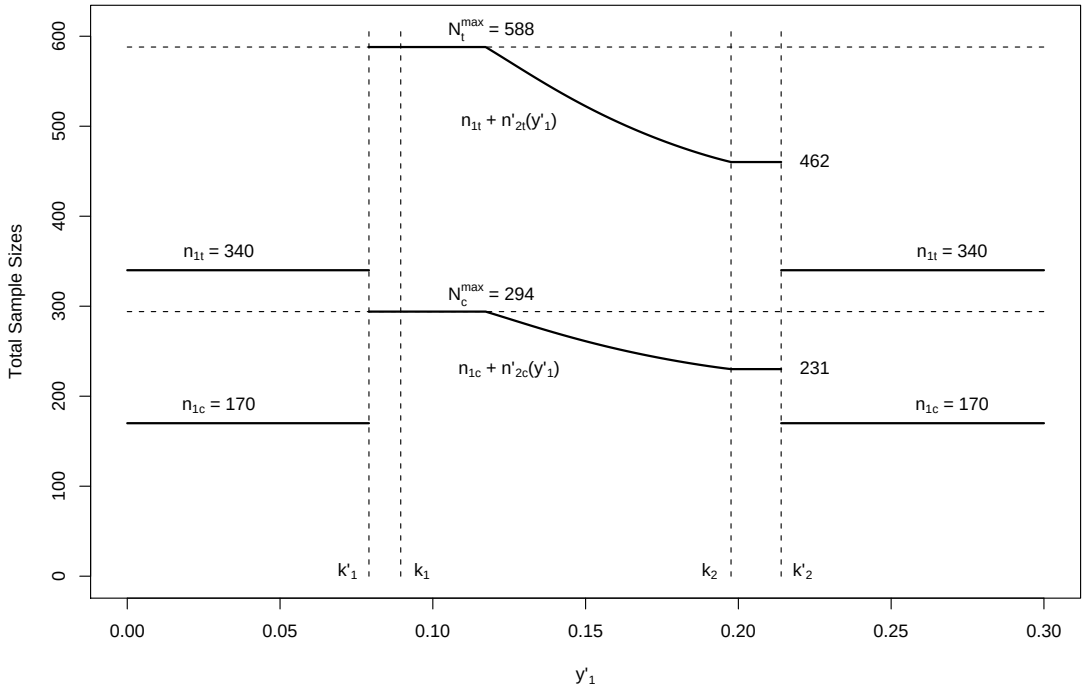


Figure 3.5: *Example of Section 3.2.5: Total Sample Sizes after accommodating missing observations in Stage I.*

for all values of $m_2'(y_1')$. For this example, $\nu_2 = 91$ and $\nu_2' = 73.9$, so $\nu_2'/\nu_2 = .81$. And we can write $m_2^*(y_1') = .81\, m_2'(y_1')$. With this $m_2^*(y_1')$, we can obtain $A'(y_1', \xi_1)$ using (3.2.28), and further calculate the post-hoc power using (3.2.29). At $\xi = \xi_1$, P[Reject in Stage I]= .508. Post-hoc power for Stage II is .382 giving the total post-hoc power of .890. The second of the assumptions is that the sample size is truncated to the observed $\nu_2' = 73.9$. Our calculation shows that at $\xi = \xi_1$, post-hoc power for Stage II is .383, therefore, the total post-hoc power is $.508 + .383 = .891$ under this assumption.

We summarize the sample sizes for this example. The one-stage conventional study's sample sizes with the same α, β and λ are $N_t = 452$ and $N_c = 226$. For this two-stage adaptive procedure, we set the maximum sample sizes to be $N_t^{max} = 588$ and $N_c^{max} = 294$ which are 30% increase from N_t and N_c. Then in Stage I, we take 75% of N_t and N_c giving $n_{1t} = 340$ and $n_{1c} = 170$. Stage II sample sizes are designed so that the maxima are below N_t^{max} and N_c^{max}. As shown in Figure 3.3, the maxima are controlled at 558 for the treatment group and 279 for the control group. In Stage I, 44 observations from the treatment group and 29 observations from the control group are missing. They account for 13% and 17% of the total number of observations within each group. Then the Stage II sample sizes are modified as shown in Figure 3.5. By this modification, the unconditional power of the procedure at the original alternative is still .90. For $y_1' = .15$, the actual sample size

for Stage II are calculated to be $\nu_{2t} = 182$ and $\nu_{2c} = 91$ from each group. In Stage II, 33 from the treatment group and 18 from the control group are missing. They account for 18% and 20% of the planned sample sizes. By adjusting the critical value, the Type I error is controlled. We also showed that the post-hoc power is .890 and .891 under the two reasonable assumptions on missingness.

3.3 Unknown Variances

Now we consider a more realistic situation where the underlying population variances are unknown. It is desirable to have reasonable estimates of the true variances, σ_t^2 and σ_c^2, at the beginning of the study, but it is often impossible to obtain good estimates. The basic idea of the approach we propose for an adaptive design is to update the estimates of the population variances at the end of Stage I, when n_{1t} and n_{1c} observations from respective groups are available to estimate the true variances. Additionally, when the Stage I sample sizes are not large enough to estimate the variances accurately using Stage I observations alone, the estimates of the variances are updated at the end of Stage II.

We introduce the following notation. Let σ_{0t}^2 and σ_{0c}^2 be the pre-study estimates. Let s_{1t}^2 and s_{1c}^2 be the estimates using the Stage I observations, as below.

$$s_{1t}^2 = \frac{\sum_{j=1}^{n_{1t}} (x_{1tj} - \bar{x}_{1t\bullet})^2}{n_{1t} - 1}, \tag{3.3.32}$$

$$s_{1c}^2 = \frac{\sum_{j=1}^{n_{1c}} (x_{1cj} - \bar{x}_{1c\bullet})^2}{n_{1c} - 1}, \tag{3.3.33}$$

where x_{1tj} is the j^{th} observation from the treatment group in Stage I, and x_{1cj} is the j^{th} observation from the control group in Stage I, and $\bar{x}_{1t\bullet}$ and $\bar{x}_{1c\bullet}$ are the Stage I means for the treatment and control group, respectively. Similarly, let s_{2t}^2 and s_{2c}^2 be the variance estimates using (only) the Stage II observations. Then, given n_{2t} and n_{2c},

$$s_{2t}^2 = \frac{\sum_{j=1}^{n_{2t}} (x_{2tj} - \bar{x}_{2t\bullet})^2}{n_{2t} - 1}, \tag{3.3.34}$$

$$s_{2c}^2 = \frac{\sum_{j=1}^{n_{2c}} (x_{2cj} - \bar{x}_{2c\bullet})^2}{n_{2c} - 1}. \tag{3.3.35}$$

We use a similar method to the one we developed for unequal sample sizes. In (3.2.15), we use a which acts like a new standard deviation for the statistic Y_1. For the argument in this section, we make an additional assumption of equal population variances; let $\sigma^2 \equiv \sigma_t^2 = \sigma_c^2$. By assuming that the underlying populations' variances are equal, we do not lose much generality. The difference between equal variance and unequal variance occurs in the definition of a and not directly in Y_1. So the extension of the method in this section to the situation in which $\sigma_t^2 \neq \sigma_c^2$ is straightforward.

The difference between these two approaches (with and without the equal population variances assumption) is essentially the difference in degrees of freedom in the estimators of the population variance. With the equal variances assumption, we use the pooled sample variance to estimate σ^2. Using Stage I data,

$$s_{1p}^2 = \frac{(n_{1t}-1)s_{1t}^2 + (n_{1c}-1)s_{1c}^2}{n_{1t}+n_{1c}-2}\,. \tag{3.3.36}$$

And from the Stage II data,

$$s_{2p}^2 = \frac{(n_{2t}-1)s_{2t}^2 + (n_{2c}-1)s_{2c}^2}{n_{2t}+n_{2c}-2}\,. \tag{3.3.37}$$

3.3.1 Motivation

An intuitive way to handle the unknown variances is to redesign Stage II using the new estimate, s_{1p}^2, as the true variance. Before revealing the Stage I means, $\bar{x}_{1t\bullet}$ and $\bar{x}_{1c\bullet}$, we would calculate s_{1p}^2 and redesign Stage II. Theoretically, it is possible to do so and still maintain the integrity of the study. Because s_{1p}^2 is independent of $\bar{x}_{1t\bullet}$ and $\bar{x}_{1c\bullet}$ theoretically in this situation, no information about $\bar{x}_{1t\bullet} - \bar{x}_{1c\bullet}$ is gained by seeing s_{1p}^2. However, practically, this is not a good solution to the problem. The variance based on all the observations can be calculated without breaking the blind. Then, as Liu (2002) argues, if the pooled variance of the two groups is revealed, the squared difference of the means of the two groups can be calculated from the available information. Thus, "partially breaking the blind" by revealing only s_{1p}^2 and using it to redesign the trial before seeing $\bar{x}_{1t\bullet}$ and $\bar{x}_{1c\bullet}$ may jeopardize the integrity of the study due to Liu's argument that s_{1p}^2 also often in essence reveals $\bar{x}_{1t\bullet} - \bar{x}_{1c\bullet}$. In Appendix B, we show in detail how $(\bar{x}_{1t\bullet} - \bar{x}_{1c\bullet})^2$ can be calculated from s_{1p}^2 in addition to the overall variance, $s_{1\bullet}^2$, without group identity.

Redesigning Stage II after breaking the blind, i.e., after revealing $\bar{x}_{1t\bullet}$, $\bar{x}_{1c\bullet}$, s_{1t}^2, s_{1c}^2, may not be done, either. As pointed out in Section 2.2.2, it is necessary to design Stage II sample size and critical values in the protocol prior to the start of Stage I. Otherwise, we cannot guarantee the Type I error rate in a rigorous point of view as shown in Liu, et.al. (2002). Stage II needs to be defined in some measurable way with respect to the Stage I observations prior to the start of the study.

Therefore, even when the population variances are unknown and they have to be estimated within the study, it is essential that the plan of Stage II be defined prior to the start of Stage I for all values of s_{1t}^2 and s_{1c}^2.

3.3.2 Ratio Of Variances

In Section 3.3.3, we propose a procedure in which the pooled estimate of the variance is updated at the end of Stage I when the sample sizes for Stage I are large. At the end of Stage II, we still use

the same estimate of the variance from Stage I. We do not use additional information from Stage II to update the estimate again. Similarly, when the Stage I sample sizes are small and the Stage II sample sizes are large, we use the variance estimate from Stage II observations only at the end of Stage II. We do not combine Stage I and Stage II observations in this case, either. It is because the variance estimate utilizing the data from only one large stage is similar to the variance estimate utilizing all the data from both stages as shown in Appendix C. When the sample size of one stage is moderate to large, the loss of information for not combining the data from two stages to estimate σ^2 is negligible. [See Appendix C.]

3.3.3 Designs With A Large Stage I Sample Sizes

Before the study begins, we use a pre-study estimate, σ_0^2, of the common population variance to design a procedure. Moreover, we assume that these estimates are the true variances for each population. Let $\lambda = n_{1c}/n_{1t}$ like before, and we define the test statistic Y_1 as follows:

$$Y_1 = \frac{\bar{X}_{1t} - \bar{X}_{1c}}{\sqrt{2}\, a}\,, \tag{3.3.38}$$

where

$$a = a(\sigma_0^2,\, \sigma_0^2,\, \lambda) = \sqrt{\frac{(1+\lambda)\sigma_0^2}{2}}\,. \tag{3.3.39}$$

Because we assume that σ_0^2 is the true variance, we have the following result for the distribution of Y_1.

$$Y_1 \sim \text{Normal}\left(\xi,\, \frac{1}{m_1}\right),$$

where $\xi = \Delta/\sqrt{2}\, a$ and $m_1 = n_{1c}$. We can use the method described in Section 3.2.2 for "planned unequal sample sizes" to design a two-stage adaptive procedure with all five specification components of Stage I and all four specification components of Stage II.

Note that if we do not make the equal variance assumption, then we use the following a in (3.3.38):

$$a = a(\sigma_{0t}^2,\, \sigma_{0c}^2,\, \lambda) = \sqrt{\frac{\lambda\sigma_{0t}^2 + \sigma_{0c}^2}{2}}\,.$$

We examine first the effect of misspecifying variance estimates if we do not make any modifications to the design at the end of Stage I. When the true variance is s_{1p}^2, the distribution of Y_1 is

$$Y_1 \sim \text{Normal}\left(\xi,\, \frac{1}{m_1}\,\frac{s_{1p}^2}{\sigma_0^2}\right).$$

The variance of Y_1 is larger than $1/m_1$ if $s_{1p}^2 > \sigma_0^2$. Then consequently, P[Reject H_0 in Stage I $|\, H_0$] $> \alpha_1$ and P[Accept H_0 in Stage I $|\, H_1$] $> \beta_1$. As a small example, the following Table 3.3 summarizes the inflation of α_1 when $\alpha_1 = .01$. The inflation of α_1, unlike the inflation of β_1, is independent of m_1 and σ_0; it only depends on the ratio of σ_0 and s_{1p}. As can be seen from this example, a

Table 3.3: *An example of inflation of α_1 as a function of s_{1p}/σ_0.*

s_{1p}/σ_0	0.5	0.75	1	1.25	1.5	1.75	2
p	.000	.001	.010	.0314	.0605	.0919	.1224

For this small example, $\alpha_1 = .01$.
p is the probability of rejecting H_0 in Stage I under H_0.

relatively small increase of s_{1p} from σ_0 results in a rather significant increase in α_1. On the other hand, we also notice that a relatively small decrease of s_{1p} results in a significant decrease in α_1. However, naturally, we are mostly concerned with inflation of α_1 and α in the case in which $\sigma_0^2 < s_{1p}^2$. Therefore, we present in this section how to modify a design when $\sigma_0^2 < s_{1p}^2$.

The most important criterion for the design is the overall Type I error rate, α. It is also desirable to keep the original power, but the necessary sample size may be enormous when s_{1p}^2 is larger than σ_0^2. In order to avoid an unrealistic solution with practically unattainable sample size, we choose the maximum sample size to be the next criterion. Let N^{max} be the pre-set maximum total sample size allowed for each group, and let $n_2^{max} = N^{max} - n_1$. If the original power is still attainable with this limit, this design controls both Type I and Type II errors. Otherwise, the power of the study will be less than $1 - \beta$.

The unknown variance problem is treated in a similar way as the missing observations problem. In the current section, Y_1 is defined in (3.3.38) which is exactly the same as (3.2.12) except for the definition of a. Let Y_1' be defined like Y_1 in (3.3.38) as follows:

$$Y_1' = \frac{\bar{X}_{1t} - \bar{X}_{1c}}{\sqrt{2}\, a}\,,$$

where $a = a(\sigma_0^2,\, \sigma_0^2,\, \lambda) = \sqrt{(1+\lambda)\sigma_0^2/2}$. The distribution of Y_1' is

$$Y_1' \sim \text{Normal}\left(\xi,\, \frac{1}{m_1'}\right), \tag{3.3.40}$$

where

$$m_1' = m_1 \frac{\sigma_0^2}{s_{1p}^2}\,. \tag{3.3.41}$$

We use the notation m_1' in (3.2.16). Although m_1' in (3.2.16) and m_1' in (3.3.41) have different meanings, we use the same notation because they are both the new "sample size" for Stage I. By

(3.3.41), it is obvious that $m_1' < m_1$ if and only if $\sigma_0^2 < s_{1p}^2$. Even when the equal population variance assumption is not made, we can still use (3.3.40) but with

$$m_1' = m_1 \frac{\lambda\sigma_{0t}^2 + \sigma_{0c}^2}{\lambda s_{1t}^2 + s_{1c}^2}\,. \tag{3.3.42}$$

For (3.3.42), the relationship of m_1' to m_1 is more complicated. Using the fact that $\lambda = n_{1c}/n_{1t}$, we can deduce that $m_1' < m_1$ if and only if $s_{1t}^2 < \sigma_{0t}^2 + \dfrac{n_{1t}}{n_{1c}}(\sigma_{0c}^2 - s_{1c}^2)$.

With Y_1' defined as in (3.3.40), we can apply the same method of redesigning Stage II as the one for the missing observation in Stage I described in Section 3.2.3. Fact 3.2.1 is easily adapted to the current situation as follows:

Fact 3.3.1. *If $s_{1p}^2 > \sigma_0^2$ then $k_1' < k_1$ and $k_2 < k_2'$. And if $s_{1p}^2 < \sigma_0^2$ then $k_1' > k_1$ and $k_2 > k_2'$.*

One difference between the current situation and the missing observations is that we do not always have $m_1' < m_1$ in the unknown variances case. It is conceivable that the prestudy estimate, σ_0^2 is a pessimistic one, and thus $s_{1p}^2 < \sigma_0^2$. As pointed out before, without making any modifications to the procedure, the Type I and II error rates are both smaller than the planned values if $s_{1p}^2 < \sigma_0^2$. However, it is possible to reduce the Stage II sample size by making modifications to the procedure when $s_{1p}^2 < \sigma_0^2$. Therefore, although the main discussion is for the case in which $\sigma_0^2 < s_{1p}^2$, we also make some comments for the case in which $s_{1p}^2 < \sigma_0^2$.

When $\sigma_0^2 < s_{1p}^2$ and thus $m_1' < m_1$, we use the new $A'(y_1', \xi_0)$ defined in (3.2.21) and the new $m_2'(y_1')$ defined in (3.2.22). When $s_{1p}^2 < \sigma_0^2$, the continuation region is narrower. Similarly to (3.2.21), we can let $A'(y_1', \xi_0) = \theta_0\, A(y_1', \xi_0)$ for $k_1' < y_1' \le k_2'$. Note that the domain is shortened and the original A function which is outside of the new domain is cut off. The constant, θ_0, for this case is greater than 1. For the Stage II sample size, $m_2'(y_1')$, the constant, η, is now some negative number, making the modified Stage II sample sizes smaller than the original ones. In the missing observation case, we use (3.2.23) to calculate an appropriate η. However, for the unknown variance case, we use a different criteria to find η. A discussion on how to find η will be given later.

Another difference between the unknown variances case and missing observations case is that even though we assume that s_{1p}^2 is the true variance, it is obviously a random variable. In the missing observations case, because we assume that observations are missing completely at random we treat n_{1c}' and n_{1t}' as constants. Because the underlying distributions of X_{1t} and X_{1c} are:

$$X_{1t} \sim \text{Normal}\left(\mu_t,\, \sigma^2\right),$$
$$X_{1c} \sim \text{Normal}\left(\mu_c,\, \sigma^2\right),$$

the distribution of s_{1p}^2 is:

$$s_{1p}^2 \sim \frac{\sigma^2}{2(n_{1t}+n_{1c}-1)}\,\chi^2_{2(n_{1t}+n_{1c}-1)}, \tag{3.3.43}$$

where χ_p^2 is a chi squared random variable with p degrees of freedom. With this new concept, one slight modification becomes possible. Instead of using s_{1p}^2 to design the new Stage II, we can find a pessimistic estimate, s_{1p}^{*2}, which is an upper bound of a confidence interval of σ^2 from Stage I. When treating s_{1p}^2 as a random variable, a concern is what effect this treatment has on the Type I error rate. However, it does not inflate the Type I error rate. The reason for this is that using the method described in this section, we can protect the Type I error rate for any value of s_{1p}^2. To show this, we define the new concept, the conditional power given s_{1p}^2, which is denoted by $\rho^*(\xi \mid s_{1p}^2)$, as

$$\rho^*(\xi \mid s_{1p}^2) = \int_{-\infty}^{\infty} A'(y_1',\, \xi) g'_\xi(y_1')\, dy_1',$$

where $g'_\xi(y_1')$ is the pdf of Y_1' assuming that the true variance is s_{1p}^2. Thus, this concept is the conditional power given s_{1p}^2. We note that for any possible value of s_{1p}^2, $A'(y_1',\, \xi_0)$ is obtained so that $\rho^*(\xi_0 \mid s_{1p}^2) = \alpha$. Therefore, regardless of the distribution of s_{1p}^2, the Type I error rate is always α. Suppose that the pdf of s_{1p}^2 is $h(s_{1p}^2)$. Then we have

$$\begin{aligned}\text{Type I error rate} &= \int_{-\infty}^{\infty} \rho^*(\xi_0 \mid s_{1p}^2) h(s_{1p}^2)\, ds_{1p}^2 \\ &= \alpha \int_{-\infty}^{\infty} h(s_{1p}^2)\, ds_{1p}^2 = \alpha.\end{aligned}$$

Thus, the unconditional Type I error rate is α.

We shift now our attention to the power at the original alternative, ξ_1. The $\rho^*(\xi \mid s_{1p}^2)$ for $\xi \neq \xi_0$ has different values for different s_{1p}^2, and it depends on η through $m_2'(y_1')$. The minimum value of $\rho^*(\xi \mid s_{1p}^2)$ for $\xi > \xi_0$ occurs when $\eta = 0$ and the maximum occurs when η is large enough so that $m_2'(y_1') = m_2^{max}$ for the entire continuation region. We can calculate this range of $\rho^*(\xi \mid s_{1p}^2)$ for any ξ. If ρ is within the range of $\rho^*(\xi_1 \mid s_{1p}^2)$, then we can find η such that the $\rho^*(\xi_1 \mid s_{1p}^2)$ is exactly equal to ρ. On the other hand, if the maximum of $\rho^*(\xi_1 \mid s_{1p}^2)$ is below ρ, then the conditional power of ρ is not attainable for the particular s_{1p}^2.

The unconditional power at the original alternative is less than $1-\beta$ because, for large values of s_{1p}^2, $\rho^*(\xi_1 \mid s_{1p}^2)$ is less than ρ unless there is essentially no limit to the Stage II sample size. The unconditional power at an alternative, ξ, depends on the true value of σ^2, and given by the form,

$$\text{Power} = \int_0^{\infty} \rho^*(\xi \mid s_{1p}^2)\, h(s_{1p}^2) ds_{1p}^2, \tag{3.3.44}$$

where $h(s_{1p}^2)$ is the pdf of s_{1p}^2, and depends on σ^2. [See (3.3.43).]

At the end of Stage II, we do not update the estimate of the variance. The key assumption in this procedure is that n_1 is large enough to treat s_{1p}^2 from Stage I as the truth. Therefore, there is no need to recalculate the post-hoc power at the end of Stage II when the variance estimate from the Stage II and the variance estimate utilizing both stages are available.

3.3.4 Example

As an example, consider the following situation. The hypotheses to be tested are $H_0 : \mu_t - \mu_c \leq 0$ against $H_1 : \mu_t - \mu_c > 0$. We set the error probabilities to be $\alpha = .025$, and $\beta = .10$ at $\mu_t - \mu_c = \Delta_1 = 1$. We assume that the underlying populations' variances are equal, but we do not assume that the common population variance, σ^2, is known. At the beginning of the study, we assume that the best available estimate for σ^2 is $\sigma_0^2 = 4^2$, and we design the trial at the start using this σ_0^2. We design a study with $\lambda = .8$, i.e., the sample size from the control group is 80% of the sample size from the treatment group. Given this λ, we can calculate, by (3.3.39), $a = \sqrt{1.8 \times 4^2/2} = 3.795$, and $\xi_1 = 1/(\sqrt{2} \times 3.795) = .1863$. As in the example of Section 3.2.5, $W = 10.5$. Thus for a one-stage conventional study, the sample sizes are $N_c = W(1+\lambda)\sigma_0^2 = 10.5 \times 1.8 \times 16 = 302$ and $N_t = 302/.8 = 378$. Suppose that the Stage I sample sizes are 80% of N_t and N_c, i.e., $n_{1t} = 302$ and $n_{1c} = 242$. The maximum sample sizes that are to be allowed are set at $N_t^{max} = 604$ and $N_c^{max} = 483$, which are each 60% increases from N_t and N_c. Further suppose that we set the Stage I error probabilities to be $\alpha_1 = .005$ and $\beta_1 = .05$. Then we can calculate Stage I critical values, $k_1 = .0806$ and $k_2 = .1656$. In summary, this particular Stage I is characterized by the following components: $\{\alpha_1 = .005, \beta_1 = .05, m_1 = 242, k_1 = .0806, k_2 = .1656\}$.

For the Stage II specification components, we choose $A(y_1, \xi_0) = .10 + 3.70(y_1 - k_1)$ and $A(y_1, \xi_1) = .8 + .885(y_1 - k_1)$. These linear equations satisfy the conditions that $\alpha = .025$ and $\rho = 1 - \beta = .90$. Using these two A-functions, we can obtain $m_2(y_1)$ and $w(y_1)$ for Stage II. This $A(y_1, \xi_1)$ is nondecreasing and $m_2(y_1)$ is nonincreasing, satisfying the conditions of Lemma 2.3.2. If σ^2 were known to be 4^2, then the design of the two-stage adaptive clinical trial is complete. We can calculate the maximum sample size, expected sample size and the power. Assuming that $\sigma^2 = 4^2$, the maximum total sample size for the control group is $m_1 + m_2(k_1) = 242 + 130 = 372$, and for the treatment group, it is $372/.8 = 465$. Stage I probabilities, the power and the expected sample sizes are summarized in Table 3.4 for various values of $\Delta \in (0, 1.25)$. However, since σ^2 is unknown, before the study begins, we need to address how to modify the study at the end of Stage I to reflect the observed s_{1p}^2.

Because $N_c^{max} = 483$, the maximum $m_2'(y_1')$ allowed is $483 - 242 = 241$. If s_{1p} were 6.0, then $\rho^*(\xi_1 \mid s_{1p}^2)$ is .90 if $m_2'(y_1') = 241$ for all $y_1' \in (k_1', k_2')$. More importantly, this implies that if $s_1 > 6.0$

Table 3.4: *Example of Section 3.3.4: Stage I probabilities, power and expected sample sizes for the original design assuming that* $\sigma^2 = \sigma_0^2 = 4^2$.

Δ	Stage I			Stage II	Power	Sample Size	
	Accept	Continue	Reject	Reject		Control	Treatment
0.00	.895	.100	.005	.020	.025	251.5	314.4
0.25	.702	.266	.032	.096	.128	266.0	332.5
0.50	.422	.448	.130	.245	.375	279.9	349.9
0.75	.179	.477	.344	.343	.687	279.9	349.9
1.00	.050	.323	.627	.273	.900	266.1	332.6
1.25	.009	.139	.853	.128	.981	251.7	314.6

The last two columns are expected total sample sizes of 2 stages for the control and treatment groups.

then $\rho^*(\xi_1 \mid s_{1p}^2) = .90$ is not attainable. For $4 < s_1 < 6.0$, we can find η so that $\rho^*(\xi_1 \mid s_{1p}^2)$ is .90. For example, if $s_1 = 4.5$ then $\eta = 59$; if $s_{1p} = 5$ then $\eta = 102$; or if $s_1 = 5.5$ then $\eta = 141$. Table 3.5

Table 3.5: *The new Stage I critical values,* θ_0 *and* η *for various* s_{1p}.

s_{1p}	k_1'	k_1	k_2	k_2'	θ_0	η
4.0	.0806	.0806	.1656	.1656	1.000	0
4.5	.0674	.0806	.1656	.1863	.6250	59
5.0	.0542	.0806	.1656	.2070	.4444	102
5.5	.0410	.0806	.1656	.2277	.3448	141
6.0	.0278	.0806	.1656	.2484	.2817	max

summarizes the new Stage I critical values, the values of θ_0, which adjust $A(y_1, \xi_0)$ function and the values of η for the cases in which the observed $s_{1p} = 4.0, 4.5, 5.0, 5.5, 6.0$.

In Figure 3.6, the original A-functions (for $\sigma = 4$) and the modified A'-functions for $s_{1p} = 5$ are shown. And in Figure 3.7, the original and the modified total sample size functions are displayed. Note that $k_1' = .0542$ and $k_2' = .2070$ when $s_{1p} = 5.0$.

To fulfill the requirement that all possible design modifications for observed s_{1p}^2 be stated prior to Stage I, we would use the following rule:

$$\text{if} \begin{cases} s_{1p} \le 6.0 & \text{then } m_2'(y_1') = min(241, m_2(y_1) + \eta) \\ & \quad \text{where } \eta \text{ is chosen so that } \rho^*(\xi_1 \mid s_{1p}^2) = .90. \\ 6.0 < s_{1p} & \text{then } m_2'(y_1') = 241 \text{ for all } y_1. \end{cases}$$

With this rule, we have a two stage adaptive design for any value of s_{1p}^2 from Stage I.

If s_{1p} is much larger than 6.0, then the $\rho^*(\xi_1 \mid s_{1p}^2)$ may be very low. For example, if $s_{1p} = 8$, then $\rho^*(\xi_1 \mid s_{1p}^2)$ is about .0.79 even we take the maximum sample size allowed in Stage II for all $y_1 \in (k_1', k_2')$. Because our original belief is $\sigma_0 = 4$, this $s_{1p} = 8$ would be twice as large as σ_0. In a conventional setting, it would lead to an increased sample size by 4-fold. Clearly, a two-stage

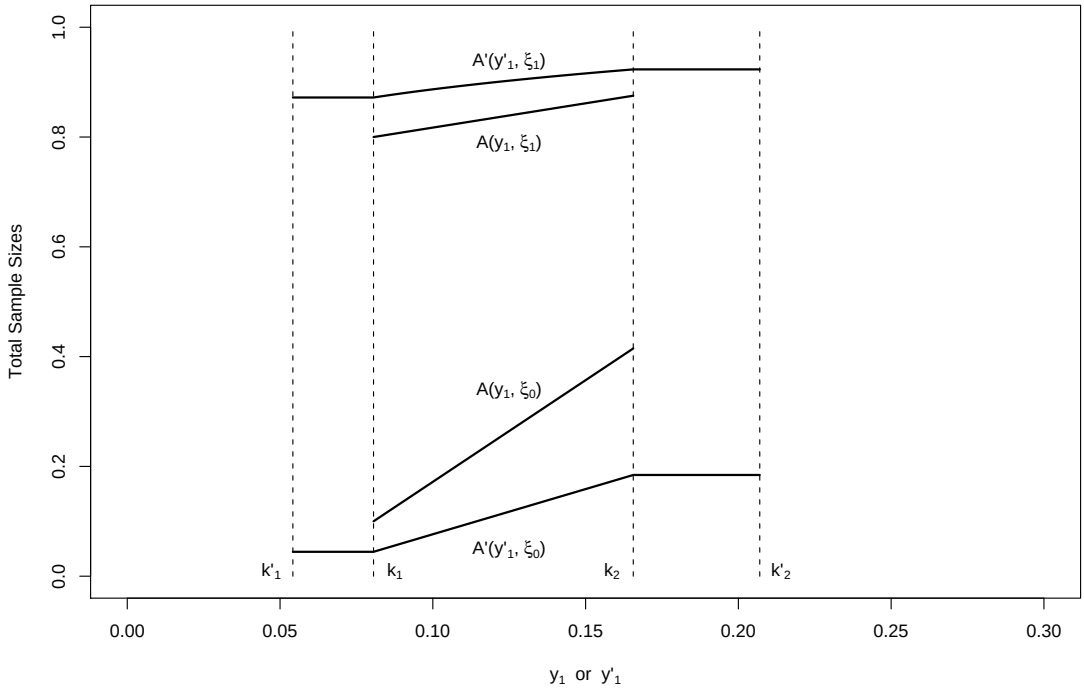


Figure 3.6: *Example of Section 3.3.4:* $A(y_1, \xi_0)$, $A(y_1, \xi_1)$, $A'(y_1', \xi_0)$, $A'(y_1', \xi_1)$.

adaptive procedure can not rescue a badly planned clinical trial, and the virtue of a careful planned study should not be underestimated.

When $\rho^*(\xi_1 \mid s_{1p}^2)$ is observed to be below an acceptable level, the clinical trial sponsor may decide not to proceed to Stage II. We do not put this formally as an additional decision rule. It should be noted that when it is decided to stop the trial for futility, α is not inflated.

In this example, if we assume the true variance is $\sigma = \sigma_0 = 4$, i.e., the pre-study guess is correct, a 99% prediction interval of s_{1p} is $(3.64, 4.37)$, which is much smaller than 6.0, the maximum s_{1p} for which the $\rho^*(\xi_1 \mid s_1^2) = .90$ is achieved. Therefore, we conclude that the unconditional power (3.3.44) is virtually equal to .90 in this example if $\sigma = 4$. We can calculate the power of this procedure for any value of σ^2 using (3.3.44).

So far in this example, we have assumed that s_{1p}^2 is the correct variance when redesigning Stage II. However, Y_2 actually does not follow the distribution with s_1^2 as its variance. The true variance is still σ^2 which remains unknown. To be conservative, we can redesign the Stage II using a pessimistic estimate of σ^2. We can find a 99% confidence interval of σ^2 and use its upperbound, s_{1p}^{*2} as the estimate of σ^2. In the current example, if we observe $s_1 = 5$, the original estimate, $\sigma_0 = 4$, is not plausible because a 99% confidence interval of σ is (4.58, 5.50). We can redesign Stage II using

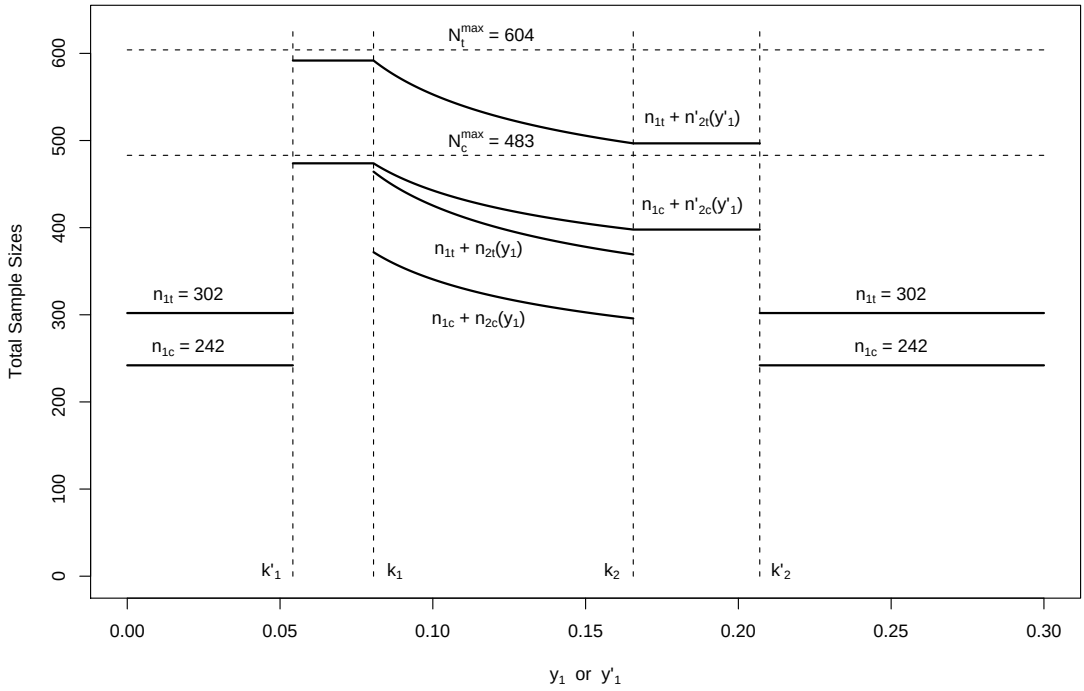


Figure 3.7: *Example of Section 3.3.4:* $n_{1t} + n_{2t}(y_1)$, $n_{1c} + n_{2c}(y_1)$, $n_{1t} + n'_{2t}(y'_1)$, $n_{1c} + n'_{2c}(y'_1)$.

s^*_{1p} which is 5.50, instead of $s_{1p} = 5.0$. When we use $s^*_{1p} = 5.5$ as the estimate of σ, $m'_2(y'_1)$ is $min(241, m_2(y'_1) + 141)$.

3.3.5 Designs With A Small Stage I Sample Sizes

In this subsection, we consider situations in which the common population variance, σ^2, is unknown, and the Stage I sample size is small. The Stage I sample size is a part of the design, and it is chosen by the investigator. However, outside factors may prevent a design from having a large Stage I. A natural remedy for such a case is to use a t distribution in place of a normal distribution. Using t distribution can protect the Type I error rate when the variance is unknown and the Stage I sample size is small. However, when σ is not known, we cannot correctly specify the Type II error rate even using a noncentral t distribution. It is not a problem particular to a two-stage adaptive procedure; the same problem exists for a conventional single-stage design. The problem exists because we do not know how far the alternative is from the null value in standardized units, and using a t distribution does not solve this problem. To calculate the Type II error rate, we need to know the distribution of the Stage I observations. The distribution is a noncentral t, and its noncentrality parameter is a function of the unknown σ.

In a two-stage adaptive procedure, the design of Stage II depends on the observations from Stage I. When Stage I sample size is small, the design of Stage II must be based on an unreliable estimate of the population variance. Even at the end of Stage II when a more reliable estimate of the population variance is available, we cannot modify the sample size. If we make a bad decision on the Stage II sample size due to an unreliable variance estimate at the end of Stage I, there is very little can be done at the end of Stage II to fix the problem. For completeness we present here a procedure, but we do not recommend using a small Stage I sample size when the population variance is unknown and must be estimated.

We assume that the underlying population variances for two groups are equal, i.e., $\sigma^2 \equiv \sigma_t^2 = \sigma_c^2$. We also assume that the sample sizes from the treatment and control groups are the same for Stage I and the same for Stage II. For Stage I, let $n_1 = n_{1t} = n_{1c}$. Furthermore, without loss of generality, assume that $\Delta_0 = 0$. Let σ_0^2 be the prestudy estimate of the variance, s_{1p}^2 be the pooled variance estimate from Stage I and s_{2p}^2 be the pooled variance estimate from Stage II. For Stage I, define

$$T_1 = \frac{\sqrt{n_1}\,(\bar{X}_{1t} - \bar{X}_{1c})}{\sqrt{2}\,s_{1p}}\,.$$

And let $h(t_1,\, \Delta,\, n_1,\, s)$ be the pdf of $t_{2(n_1-1)}\left(\dfrac{\sqrt{n_1}\,\Delta}{\sqrt{2}\,s}\right)$, where $t_v(\delta)$ represents a non-central t distribution with degrees of freedom, v, and the non-centrality parameter, δ. The variable s appears in the non-centrality parameter, and $s = \sigma$ gives the true distribution of T_1. However, because σ is unknown, we use an estimate for it such as σ_0 or s_{1p}, where s denotes one of these quantities.

Let $n_2(t_1, s)$ be the Stage II sample size function. The Stage II sample size is a function of the stage I observations, t_1, and the estimate, s, of the standard deviation. Given $T_1 = t_1$ let

$$T_2 = \frac{\sqrt{n_2(t_1,s)}\,(\bar{X}_{2t} - \bar{X}_{2c})}{\sqrt{2}\,s_{2p}}\,.$$

Table 3.6: *Example of Section 3.3.4: Stage I probabilities, power and expected sample sizes for the modified design for* $s_{1p} = 5$.

μ	Stage I			Stage II	Power	Sample Size	
	Accept	Continue	Reject	Reject		Control	Treatment
0.00	.750	.245	.005	.020	.025	292.2	365.3
0.25	.538	.439	.023	.107	.130	328.9	411.1
0.50	.314	.608	.078	.303	.381	357.4	446.8
0.75	.143	.655	.201	.491	.692	361.1	451.4
1.00	.050	.551	.399	.501	.900	338.3	422.9
1.25	.013	.360	.627	.351	.978	302.8	378.5

The last two columns are expected total sample sizes of 2 stages for the control and treatment groups.

Let $G(t_1,\, n_2(t_1,\, s),\, \Delta,\, s)$ be the cdf of $t_{2(n_2(t_1,\, s)-1)}\left(\dfrac{\sqrt{n_2(t_1,\, s)}\,\Delta}{\sqrt{2}\,s}\right)$.

Stage I decision rule is

$$\text{if}\begin{cases} T_1 \le k_2 & \text{continue to Stage II,} \\ k_2 < T_1 & \text{stop and reject } H_0. \end{cases}$$

Note that we do not allow "stop and accept H_0" in Stage I. The reason is that we cannot control Type II error rate for Stage I when the sample size is small. Type I error rate is protected by choosing k_2 that satisfies $\alpha_1 = P[T_1 > k_2]$ where $T_1 \sim t_{2(n_1-1)}$. Because the distribution of T_1 does not depend on variance estimate under H_0, α_1 is correct no matter what s_{1p}^2 is. Using different notation, we have $\alpha_1 = \int_{k_2}^{\infty} h(t_1,\, 0,\, \bullet)\, dt_1$. The third argument of h is irrelevant under H_0 because T_1 has a central t distribution.

The specification components of Stage I are $\{\alpha_1,\, n_1,\, k_2\}$. When either α_1 or k_2 is specified in addition to n_1, we have all the specification components for the Stage I.

For Stage II, we need to find sample size and critical value. The sample size is, as mentioned before, $n_2(t_1,\, s)$. The critical value is also a function of T_1 and s; let $w(t_1,\, s)$ denote the critical value. The Stage II decision rule is

$$\text{if}\begin{cases} T_2 \le w(t_1,\, s) & \text{accept } H_0, \\ w(t_1,\, s) < T_2 & \text{reject } H_0. \end{cases} \tag{3.3.45}$$

In order to specify $n_2(t_1,\, s)$ and $w(t_1,\, s)$ so that the Type I error rate and power are controlled, we define the conditional power function $A(t_1,\, \Delta,\, s)$ as follows:

$$A(t_1,\, \Delta,\, s) = P[\,\text{Reject } H_0 \text{ in Stage II}\,|\,T_1 = t_1,\, \Delta,\, s\,].$$

It follows that

$$\begin{aligned} A(t_1,\, \Delta,\, s,) &= P[\,T_2 > w(t_1,\, s)\,|\,t_1,\, \Delta,\, s\,] \\ &= 1 - G\,(w(t_1,\, s), n_2(t_1,\, s),\, \Delta,\, s)\,. \end{aligned}$$

Similarly to the discussion in Chapter 2, the specification components for Stage II are $\{A(t_1,\, 0,\, s)$, $A(t_1,\, \Delta_1,\, s)$, $n_2(t_1,\, s)$ and $w(t_1,\, s)\}$. In the planning before the Stage I, we choose 2 of these 4 components so that

$$\alpha_2 = \int_{-\infty}^{k_2} A(t_1,\, 0,\, \sigma_0) h(t_1,\, 0,\, \bullet)\, dt_1, \tag{3.3.46}$$

$$\rho_2 = \int_{-\infty}^{k_2} A(t_1,\, \Delta_1,\, \sigma_0) h(t_1,\, \Delta_1,\, \sigma_0)\, dt_1. \tag{3.3.47}$$

Under H_0, (3.3.46) gives the correct α_2. Under H_1, the correct distribution of T_1 is $h(t_1,\, \Delta_1,\, \sigma)$, but since σ is unknown, we replace it with the best estimator available, which is σ_0 at this point. So in the planning before the Stage I, the four components are:

$$\{A(t_1,\, 0,\, \sigma_0),\, A(t_1,\, \Delta_1,\, \sigma_0),\, n_2(t_1, \sigma_0),\, w(t_1, \sigma_0)\}.$$

Then we need to specify how to update these components including the sample size and critical value of Stage II. At the end of Stage I, we have $\bar{x}_{1t}$, $\bar{x}_{1c}$ and s_{1p}^2. Therefore we can calculate t_1. There is no need to change the Stage I critical value, k_2, because it is still true that

$$\alpha_2 = \int_{k_2}^{\infty} h(t_1,\, 0,\, \bullet)\, dt_1.$$

The correct distribution of T_1 under H_0 is independent of the Stage I variance estimate.

Now we consider updating Stage II specification components. There is no need to change $A(t_1,\, 0,\, \sigma_0)$ because (3.3.46) is independent of the Stage I variance estimate. For the power requirement similar to (3.3.47), we need $A(t_1,\, \Delta_1,\, s_{1p})$ such that

$$\rho_2 = \int_{-\infty}^{k_2} A(t_1,\, \Delta_1,\, s_{1p}) h(t_1,\, \Delta_1,\, s_{1p})\, dt_1. \tag{3.3.48}$$

Because controlling Stage II sample size is often of interest, we may choose to specify $n_2(t_1,\, s_{1p}) = n_2(t_1,\, \sigma_0) + \eta$ so that (3.3.48) is true or so that $n_2(t_1,\, s_{1p}) = \max(n_2)$ for all $t_1 \in (-\infty,\, k_2)$. the new critical value function is then $w(t_1,\, s_{1p})$. Note that 3.3.48 does not give the correct Stage II power because the distribution of T_1 is $h(t_1,\, \Delta_1,\, \sigma)$, but 3.3.48 uses the "best guess" for the distribution of T_1 at this point.

Finally at the end of Stage II, we obtain s_{2p}^2, which we treat as a reliable estimate of σ^2. The conditional Type I error rate, i.e., $A(t_1,\, 0,\, \sigma_0)$ should not be changed because it does not depend on the variance estimate. Type I error rate is correct without any modification to $A(t_1,\, 0,\, \sigma_0)$. Moreover, the sample size cannot be changed because the sample has already been taken. So we have $n_2(t_1,\, s_{2p}) = n_2(t_1,\, s_{1p})$. Because two out of four specification components are fixed, the remaining two cannot be modified, either. We have $A(t_1,\, \Delta_1,\, s_{2p}) = A(t_1,\, \Delta_1,\, s_{1p})$ and $w(t_1,\, s_{2p}) = w(t_1,\, s_{1p})$. Then we can compute a Stage II "posterior" power by:

$$\rho_2 = \int_{-\infty}^{k_2} A(t_1,\, \Delta_1,\, s_{2p}) h(t_1,\, \Delta_1,\, s_{2p})\, dt_1.$$

3.4 Discussion

In this chapter, we discuss two practically important extensions of the framework for two-stage adaptive procedures. The first of the two extension, unequal sample sizes, is categorized into planned

unequal sample sizes and unequal sample sizes due to missing observations. They are treated by defining a which acts like σ, and m_1 which acts like n_1. The original framework is applicable with these minor changes. For missing observations, we only consider MCAR, missing completely at random. When there are some missing observations in Stage I, the critical values, k_1 and k_2 are updated to keep α_1 and β_1 as planned. We show that the continuation region is wider when there are missing observations in Stage I. Hence, the Stage II specification components need to be extended. We provide one method which controls the maximum sample sizes in addition to the Type I error rate. The method we present is not the only one, and one can develop similar procedures to accommodate missing observations in Stage I. The main reason that we choose this particular one method is that it controls two of the more important aspects of a design, namely Type I error probability and the sample size. As a result, the desired power of the study may not be achievable. However, in order to achieve a particular power while keeping the Type I error rate fixed, the necessary sample sizes may be enormous. We want to avoid developing a procedure which may require unpractical and unattainable sample sizes.

When there are some missing observations in Stage II, there is little that we can do beyond protecting the Type I error rate. Because the samples have already been taken, there is no concept of adjusting the sample size. Type I error rate is kept at the targeted level simply by keeping $A(y_1, \xi_0)$ unchanged at the end of Stage II. Then we have the connections among 4 specification components, and we can calculate the new critical value functions and the conditional power function at any alternative. The post-hoc power can be calculated, but we recognize that the post-hoc power is dependent on the assumptions one makes on missingness. In other words, we have to assess what the missingness would be for other values of y_1 than the actual observed one. Because this assumptions is not verifiable, we do not recommend using the post-hoc power, although we present how it is computed for the completeness of the discussion.

As mentioned in Chapter 2, we view it necessary to specify the plan of Stage II before Stage I begins. The same argument is applicable to the situations with missing observations. It is necessary to specify before Stage I begins all the actions at the end of Stage I with regard to potential patterns of missingness. In other words, we cannot choose how to modify the A-functions after seeing the Stage I observations, or after finding out how many observation are missing in Stage I. The method presented in this chapter uniquely determines all the Stage II specification components for any missing- observations behavior in Stage I.

The second extension is to unknown variances. For simplicity, we develop the original framework assuming the population variance for each of the treatment and control groups is known. In this section, we present an extension to the original framework to handle unknown variances. The problem

is divided into two parts. The first part is when the Stage I sample sizes are large, and the second part is when the Stage I sample sizes are small and the Stage II sample sizes are large. Our main focus is on the first situation. When Stage I is sufficiently large and a variance estimate from Stage I is reliable, we can reflect the estimate from Stage I in the design of Stage II. The method we present for this situation is similar to the one used for the missing observations in Stage I. We update m_1 at the end of Stage I to m_1' which reflects the new estimate of the population variance.

If the Stage I sample size is adequate, we do not update the variance estimate again at the end of Stage II. Not using a variance estimate utilizing all the data from Stage I and Stage II makes the mathematics of the procedure simpler. In Appendix C, we study the relationship between a variance estimate from Stage I and a variance estimate using the entire data.

There are many additional problems and concerns when the population variance is unknown and the Stage I sample sizes are small. We presented a procedure in this chapter, but one should be cautious in considering applying this procedure. Because the design of Stage I is based on a prestudy estimate of the variance and a reliable estimate is not available at the end of Stage I, the design of Stage II must be based on an unreliable estimate of the variance. When a good estimate of the variance becomes available at the end of Stage II, there is little we can do to "fix" the design. An additional issue arises when the Stage I estimate of the variance is very different from the prestudy estimate. One would need a rule to update the pre-study estimate by the Stage I estimate, but it is very difficult to see how to establish such a rule. The rule would depend on how much confidence one has on the prestudy estimate. The question is which of these two estimates or their combination are to be used in designing Stage II. For example, suppose that we start with a prestudy estimate, $\sigma_0 = 4$, and in Stage I with a small sample size, we observe $s_{1p} = 6$. Then we could design Stage II using either σ_0 or s_{1p} or a combination. If we believe that σ_0 is close to the truth than the small sample estimator from Stage I, we should use σ_0. Or, if we have little confidence in σ_0, we can update using s_{1p}. In addition, we can obtain a confidence interval of σ using s_{1p} and use its upper bound as a pessimistic estimate. We do want to emphasize that for small sample sizes in Stage I, the Type I error rate is protected using the procedure given in this chapter.

Chapter 4

Switching Study Objectives Between Noninferiority And Superiority

4.1 Introduction

In a clinical trial setting, the hypotheses of interest are often ones of noninferiority or of superiority. The objective of a noninferiority trial is to show that the new treatment is at least as good as the active control, while the objective of a superiority trial is to show that the new treatment is better than the control.

In a noninferiority trial, the null hypothesis, $H_0^{NI} : \mu_t - \mu_c \leq -\Delta_I$ is tested against the alternative hypothesis, $H_1^{NI} : \mu_t - \mu_c > -\Delta_I$, where Δ_I is a preset positive value (noninferiority margin). In a superiority trial, the null hypothesis, $H_0^S : \mu_t - \mu_c \leq 0$ is tested against $H_1^S : \mu_t - \mu_c > 0$. Customarily, the power of a noninferiority trial is set at the alternative $\Delta = 0$, but in some situations one may want to power the study at some value other than 0. Therefore, we let Δ_N to be the point at which the power of the noninferiority trial is set. The power of the superiority trial is set at some positive Δ_S.

Noninferiority and superiority can be tested in the same trial without multiplicity penalties. This is because the two null hypotheses can be viewed as nested hypotheses. [See Aras (2001), Morikawa and Yoshida (1995) and Marcus, Peritz and Gabriel (1976).] Because of the lack of a multiplicity penalty, as long as Δ_I is prespecified, these two hypotheses can always be tested simultaneously at the end of an active control study.

In a context of noninferiority trial, the lack of assay sensitivity is well documented. [e.g., Hwang and Morikawa (1999), Snapinn (2000), Aras (2001)] Assay sensitivity is the ability to demonstrate a difference between treatments when such a difference truly exists. In establishing superiority of a test treatment, assay sensitivity is automatically established. However, the same is not true in a noninferiority trial. A poorly executed noninferiority trial biases toward "no difference" and increases the likelihood that an ineffective treatment be found noninferior to an efficacious control. Therefore,

in a noninferiority trial assay sensitivity must be demonstrated separately. Thus, for the case in which the intent of the trial is to show superiority but only the noninferiority conclusion is reached, CPMP (2000) cautions about the lack of assay sensitivity. In a two-stage adaptive procedure, even when the trial is started with superiority as its main study objective, because there is a chance of showing only noninferiority, demonstrating assay sensitivity should be a consideration from the beginning.

When a clinical trial is planned for the main design objective (noninferiority or superiority) and the hypotheses for the other objective are also tested at the end of a single stage conventional study, the power of the second test is not controlled. It may be either too high or too low depending on the relation of $\Delta_N - \Delta_I$ and Δ_S. For example, suppose that the main objective is to establish the noninferiority of the treatment in question to a competitor and the noninferiority margin is set at $\Delta_I = -1$. The sample size, N_I, from each of the control and treatment groups of this single-stage study is planned so that $\alpha = .025$ and the power at $\mu_t - \mu_c = \Delta_N = 0$ is .90. Suppose that, if one wishes to presuppose superiority, the power would be evaluated at $\mu_t - \mu_c = \Delta_s = .5$. Because $\Delta_S - 0$ is only half of $\Delta_N - \Delta_I$, if the primary objective were superiority and the sample size were planned for that objective, then the sample size, N_S, would be four times as large as N_I, the sample size of a noninferiority trial with the same Type I and Type II error rates. However, suppose that we were to test these two sets of hypotheses in one trial for which the design objective is noninferiority, and let N_I be the sample size for an $\alpha = .025$, 90% power test, then the power of the superiority test would be only 36.8%.

For this situation in which both superiority and noninferiority hypotheses are considered, a two-stage adaptive procedure can be advantageous because a two-stage adaptive procedure allows changes to be made at the end of Stage I. We propose that the primary design objective be chosen and the Stage II sample sizes and critical values be adjusted based on the data accumulated in Stage I so that the power of both hypotheses are controlled.

There are some complications that arise for two-stage adaptive procedures which allow for the possibility of switching the main study objectives. These complications can be classified into two categories: complications due to the fact that there are two sets of hypotheses, and complications due to the adaptive nature of the trial.

The first category includes issues regarding the power functions. In a single-stage setting with only one set of hypotheses, the power function is defined as the probability of rejecting the null hypothesis as a function of Δ. However, for the current situation, there are two null hypotheses, and consequently we need to define two separate power functions. The first one is the probability of rejecting H_0^{NI}, and the second one is the probability of rejecting H_0^S. Because H_0^S is nested within

H_0^{NI}, rejecting H_0^{NI} implies that H_0^S is also rejected. Therefore, the power function associated with H_0^{NI} is always larger than the power function associated with H_0^S.

The issues concerning Type I and Type II errors are related to the preceding ones concerning power functions. When there are two sets of hypotheses, there are more than the two possible conclusions in a single-stage trial: accepting null and rejecting null. In Section 4.2, we describe the types of errors that need to be controlled for an adaptive procedures in which the possible switching of study objectives are allowed.

Issues that arise due to the adaptive nature of the procedures are, again, related to the power functions. To control Type I error rate for a composite null hypothesis, the power function needs to be bounded above by α for all the Δ's such that $\Delta \leq \Delta_I$ for the noninferiority hypotheses. The monotonicity of the power function is an easy way to guarantee that. It is shown in Chapter 2 that even without a possible switch of the study objectives, it is not a simple task to guarantee the monotonicity of the power function in an adaptive procedure. With the additional option to switch the main study objective at the end of Stage I, the monotonicity of the power function becomes even more difficult to analytically prove. A discussion on the monotonicity of power functions is given in Section 4.6.

A related issue is the natural ordering of the decisions at the end of Stage I. It makes intuitive sense that the decisions at the end of Stage I are ordered as:

- "conclude inferiority and stop,"
- "proceed to Stage II with noninferiority as the main objective (Stage II-N),"
- "conclude noninferiority and stop,"
- "proceed to Stage II with superiority as the main objective (Stage II-S),"
- "conclude superiority and stop".

A question is whether to allow concluding "inferiority" in Stage II-S, and whether to allow concluding "superiority" in Stage II-N. According to the ordering of the decisions mentioned above, "proceeding to Stage II-S" gives more evidence in Stage I against inferiority than "concluding noninferiority and stop". Does "proceeding to Stage II-S" automatically mean at least"concluding noninferiority"? And does "proceeding to Stage II-N" automatically mean at least "concluding not superiority"? Our framework, which is established for the most general case where all decisions are possible at the end of Stage I and at the end of Stage II, can produce procedures with various design characteristics, and thus we have the possibility to design a procedure that does not conclude noninferiority in Stage II-S or that does not conclude superiority in Stage II-N.

4.2 Types Of Errors

Recall that there are the following two sets of hypotheses.

$$H_0^{NI} : \mu_t - \mu_c \leq -\Delta_I, \qquad H_0^S : \mu_t - \mu_c \leq 0,$$
$$H_1^{NI} : \mu_t - \mu_c > -\Delta_I, \qquad H_1^S : \mu_t - \mu_c > 0,$$

where $\Delta_I > 0$ is prespecified. When we test both sets of hypotheses, there are three different conclusions we can reach at the end of the study. The first one is to accept H_0^{NI}, i.e., to conclude inferiority. The second one is to reject H_0^{NI} and accept H_0^S, i.e., to conclude noninferiority but not superiority. The last one is to reject H_0^S, i.e., to conclude superiority. The fact that there are only three conclusions is a special feature of combining noninferiority and superiority objectives, that stems from the fact that these null hypotheses are nested hypotheses.

In order to make notation consistent with Chapter 2, we use ξ instead of μ and Δ. In general, let $\xi = (\mu_t - \mu_c)/\sqrt{2}\,\sigma$. Some special values are: $\xi_I \equiv \Delta_I/\sqrt{2}\,\sigma$, $\xi_N \equiv \Delta_N/\sqrt{2}\,\sigma$ and $\xi_S \equiv \Delta_S/\sqrt{2}\,\sigma$.

We can regard this problem as a three-decision problem by defining the following three disjoint sets: $I = \{\xi : \xi \leq \xi_I\}$, $N = \{\xi : \xi_I < \xi \leq 0\}$ and $S = \{\xi : 0 < \xi\}$. Then these three sets correspond to the three decisions mentioned above. The set I is equivalent to accepting H_0^{NI}, N is equivalent to rejecting H_0^{NI} and accepting H_0^S, and S is equivalent to rejecting H_0^S. Note that I, N and S span the entire parameter space, $\Re$. Only one of these three states is true, and choosing one of the other two results in an error. When the truth is I, concluding N or S are Type I errors because we reject H_0^{NI} when it is true. When the truth is N, concluding S is a Type I error because we reject H_0^S when it is the true state, and concluding I is a Type II error because we fail to reject H_0^{NI}. When the truth is S, concluding N by failing to reject H_0^S and concluding I by failing to reject H_0^S and H_1^{NI} are both Type II errors.

When we consider the two study objectives, noninferiority and superiority, there are two sets of hypotheses, and consequently, there are two power functions. They are:

$$\psi_{I^c}(\xi) = P_\xi[\text{Reject } H_0^{NI}] = P_\xi[\text{Conclude } N \text{ or } S], \tag{4.2.1}$$
$$\psi_S(\xi) = P_\xi[\text{Reject } H_0^S] = P_\xi[\text{Conclude } S]. \tag{4.2.2}$$

The subscript of the first power function is I^c because "Conclude N or S" is the same as "conclude I^c" . The first one is the power function for the noninferiority objective, and the second one is the power function for the superiority objective. When the true state is I, the Type I error rate, denoted by α_I, is

$$\alpha_I = \sup_{\xi \leq \xi_I} \psi_{I^c}(\xi),$$

and when the true state is N, the Type I error rate, denoted by α_N, is

$$\alpha_N = \sup_{\xi \leq 0} \psi_S(\xi).$$

As mentioned in Section 4.1, it is not easy to guarantee the monotonicity of these power functions for adaptive designs. In Section 4.6, the monotonicity of the power functions are discussed in detail, but in this section, we simply assume that these power functions are monotonically nondecreasing. Using the monotonicity, we conclude that

$$\alpha_I = \psi_{I^c}(\xi_I), \tag{4.2.3}$$

$$\alpha_N = \psi_S(0). \tag{4.2.4}$$

Controlling α_I and α_N satisfies the requirements on the Type I error rate. However, it may be additionally desirable to control $P[\text{Conclude } S \,|\, \xi \leq \xi_I]$, because it is the probability of the error which occurs when we conclude superiority of the treatment in question when the truth is actually an inferior treatment. This represents a major risk to consumers, in that the new treatment is Δ_I worse than the standard and yet we would claim its superiority. By the monotonicity of $\psi_S(\xi)$, we have

$$\alpha_{IS} \equiv \sup_{\xi \leq \xi_I} \psi_S(\xi) = \psi_S(\xi_I).$$

The subscript "IS" means that the true state is I and we conclude S. In general, when there are two subscripts are attached to α's and β's, the first one represents the true state and the second one represents the conclusion. When there is only one subscript, it represents the true state, and the second one is omitted because either the conclusion is not specified or there is only one conclusion for the particular power.

So far we have considered α_I, α_N and α_{IS}. When the true state is I, one can commit a Type I error by concluding either N or S. We have considered the error "IS" individually in α_{IS} but not the error "IN" yet. So define α_{IN} as follows:

$$\alpha_{IN} = \alpha_I - \alpha_{IS}. \tag{4.2.5}$$

To analyze this probability more closely, define another power function, $\psi_N(\xi)$, as

$$\psi_N(\xi) = P_\xi[\text{Conclude } N].$$

Then from (4.2.1) and (4.2.2), $\psi_N(\xi) = \psi_{I^c}(\xi) - \psi_S(\xi)$. So $\alpha_{IN} = \psi_N(\xi_I)$. Unlike $\psi_{I^c}(\xi)$ and $\psi_S(\xi)$, the supremum of $\psi_N(\xi)$ in $\xi \leq \xi_I$ does not occur at the border, $\xi = \xi_I$. The shape of $\psi_N(\xi)$ is not

easy to analyze because $\psi_N(\xi)$ is the difference of two nondecreasing functions. Therefore, although we use α_{IN} to represent the probability of the error "IN," it is not the supremum of $\psi_N(\xi)$ in $\xi \leq \xi_I$. However, obviously by (4.2.5), α_{IN} is bounded above by α_I.

In the rest of this chapter, when we refer to α's, as well as β's and ρ's, we indicate in parenthesis at which value the power functions are evaluated. The four error rates we have considered so far are $\alpha_I(\xi_I)$, $\alpha_N(0)$, $\alpha_{IS}(\xi_I)$, $\alpha_{IN}(\xi_I)$.

Now we consider Type II errors. For the noninferiority objective, the Type II error probability is $\beta_N(\xi_N) \equiv 1 - \psi_{I^c}(\xi_N)$. For the superiority objective, the Type II error probability is $\beta_S(\xi_S) \equiv 1 - \psi_S(\xi_S)$. Additionally, it may be desirable to control $\beta_{SI}(\xi_S) \equiv 1 - \psi_{I^c}(\xi_S)$ individually because it is the probability of the error that causes a great loss for the clinical trial sponsor, who actually has a superior treatment to a competitor but fails to even claim noninferiority to the competitor. And we define $\beta_{SN}(\xi_S) \equiv \beta_S(\xi_S) - \beta_{SI}(\xi_S)$, which is bounded above by $\beta_S(\xi_S)$.

Table 4.1 summarizes the probabilities of various Type I and Type II errors and their representations using the power functions. As can be seen from Table 4.1, under I, the two power functions

Table 4.1: *Probabilities of Type I and Type II errors and their representations in terms of the power functions.*

<table>
<tr><td rowspan="2"></td><td rowspan="2"></td><td colspan="3">Conclusion</td></tr>
<tr><td>I</td><td>N</td><td>S</td></tr>
<tr><td rowspan="6">Truth</td><td rowspan="2">I</td><td></td><td colspan="2">$\alpha_I(\xi_I) = \psi_{I^c}(\xi_I)$</td></tr>
<tr><td></td><td>$\alpha_{IN}(\xi_I) = \psi_N(\xi_I)$</td><td>$\alpha_{IS}(\xi_I) = \psi_S(\xi_I)$</td></tr>
<tr><td rowspan="2">N</td><td>$\beta_N(\xi_N) = 1 - \psi_{I^c}(\xi_N)$</td><td colspan="2"></td></tr>
<tr><td colspan="2"></td><td>$\alpha_N(0) = \psi_S(0)$</td></tr>
<tr><td rowspan="2">S</td><td>$\beta_{SI}(\xi_S) = 1 - \psi_{I^c}(\xi_S)$</td><td>$\beta_{SN}(\xi_S) = \psi_N(\xi_S)$</td><td></td></tr>
<tr><td colspan="2">$\beta_S(\xi_S) = 1 - \psi_S(\xi_S)$</td><td></td></tr>
</table>

are evaluated at the same point, ξ_I, for $\alpha_I(\xi_I)$ and $\alpha_{IS}(\xi_I)$. Similarly, under S, the two power functions are evaluated at ξ_S to obtain both $\beta_{SI}(\xi_S)$ and $\beta_S(\xi_S)$. In contrast, under N, the two power functions are evaluated at different ξ's for $\alpha_N(\xi_N)$ and $\beta_N(0)$ unless $\xi_N = 0$.

In order to satisfy the requirements on the Type I error we set $\alpha_I(\xi_I) = \alpha$ and $\alpha_N(0) = \alpha$. Because the states I and N cannot be true simultaneously, we can set these two Type I error probabilities individually at α. The Type II error probabilities for the noninferiority objective and superiority objective can be set at different values.

4.3 Framework

4.3.1 Stage I

Consider a two-stage adaptive design where in Stage I, n_1 observations are taken from each of the treatment and control groups. The underlying distributions are independent normals with means μ_t and μ_c and variances σ_t^2 and σ_c^2. So we have

$$\bar{X}_{1t} \sim \text{Normal}\left(\mu_t, \frac{\sigma_t^2}{n_1}\right)$$
$$\bar{X}_{1c} \sim \text{Normal}\left(\mu_c, \frac{\sigma_c^2}{n_1}\right).$$

We assume that both σ_t^2 and σ_c^2 are known. Define

$$Y_1 = \frac{\bar{X}_{1t} - \bar{X}_{1c}}{\sqrt{2}\,\sigma},$$

where $\sigma^2 = (\sigma_t^2 + \sigma_c^2)/2$. Note that Y_1 is a test statistic based upon unblinded data. The distribution of Y_1 is

$$Y_1 \sim \text{Normal}\left(\xi, \frac{1}{n_1}\right),$$

where $\xi = \dfrac{\mu_t - \mu_c}{\sqrt{2}\,\sigma}$. At the end of Stage I, based on the observed y_1, a decision is made according to the following rule.

$$\text{If } \begin{cases} Y_1 \in (-\infty, k_1^I] & \text{stop and conclude } I, \\ Y_1 \in (k_1^I, k_2^I] & \text{continue to Stage II-N with primary objective of noninferiority,} \\ Y_1 \in (k_2^I, k_1^S] & \text{stop and conclude } N, \\ Y_1 \in (k_1^S, k_2^S] & \text{continue to Stage II-S with primary objective of superiority,} \\ Y_1 \in (k_2^S, \infty) & \text{stop and conclude } S. \end{cases}$$

We have $k_1^I \le k_2^I \le k_1^S \le k_2^S$ with at least one inequality holding. This rule is most general when all the k-values (critical values) are distinct and different from $\pm\infty$. Various special cases are possible by choosing specific k-values, and they are briefly discussed in Section 5.3.

4.3.2 Stage II-N

In Stage II-N, $n_2(y_1)$ additional samples are taken from each of the control and treatment groups. The sample size, $n_2(y_1)$, is allowed to be a function of y_1, and the method of choosing $n_2(y_1)$ is discussed in Section 4.4. Let $\bar{X}_{2t}$ and $\bar{X}_{2c}$ be the means of the observations in Stage II-N from the treatment and control groups, respectively. Define

$$Y_2 = \frac{\bar{X}_{2t} - \bar{X}_{2c}}{\sqrt{2}\,\sigma}. \tag{4.3.6}$$

The Stage I observations influence the distribution of Y_2 only through the dependence of the sample size $n_2(y_1)$ on y_1. Thus, the conditional distribution of Y_2 given that $Y_1 = y_1$ is Normal $(\xi,\ 1/n_2(y_1))$. At the end of Stage II-N, the decision is made according to the following criteria.

$$\text{If } \begin{cases} Y_2 \in (-\infty, w_1(y_1)] & \text{stop and conclude } I, \\ Y_2 \in (w_1(y_1), w_2(y_1)] & \text{stop and conclude } N, \\ Y_2 \in (w_2(y_1), \infty) & \text{stop and conclude } S. \end{cases}$$

Like $n_2(y_1)$, the critical values, $w_1(y_1)$ and $w_2(y_1)$, are allowed to be functions of y_1. It is possible to conclude H_S even in Stage II-N. However, we can make P[Conclude S in Stage II-N]$= 0$ by choosing $w_2(y_1) = \infty$ for all y_1.

4.3.3 Stage II-S

In Stage II-S, $m_2(y_1)$ additional samples are taken from each of the control and treatment groups. Let Y_2 be defined as in (4.3.6). The distribution of Y_2 given $Y_1 = y_1$ is Normal $(\xi,\ 1/m_2(y_1))$. At the end of Stage II-S,

$$\text{if } \begin{cases} Y_2 \in (-\infty, v_1(y_1)] & \text{stop and conclude } I, \\ Y_2 \in (v_1(y_1), v_2(y_1)] & \text{stop and conclude } N, \\ Y_2 \in (v_2(y_1), \infty) & \text{stop and conclude } S. \end{cases}$$

We have two critical value functions for Stage II-S also. Unless $v_1(y_1) = -\infty$ for all y_1, there is nonzero probability of concluding I in Stage II-S.

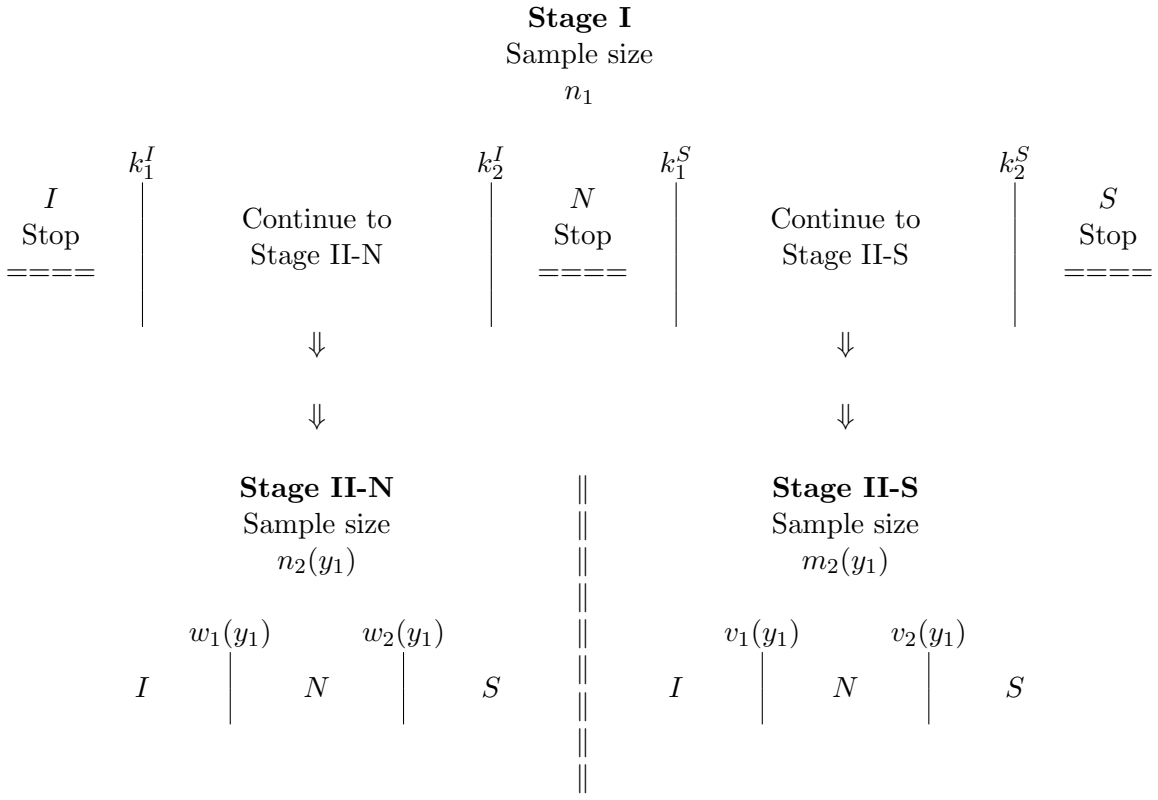
In summary, in order to define a two-stage adaptive procedure with a possible switching of hypotheses, we need n_1, k_1^I, k_2^I, k_1^S and k_2^S for Stage I; $n_2(y_1)$, $w_1(y_1)$ and $w_2(y_1)$ for Stage II-N for $y_1 \in (k_1^I, k_2^I]$; $m_2(y_1)$, $v_1(y_1)$ and $v_2(y_1)$ for Stage II-S for $y_1 \in (k_1^S, k_2^S\,]$. Both $n_2(y_1)$ and $m_2(y_1)$ are the sample size functions for Stage II, and they can be regarded as a single function evaluated at different intervals of y_1. In other words, we can define a single function in the entire range of y_1 that takes the same values as $n_2(y_1)$ for y_1 in $(k_1^I,\ k_2^I]$ and the same values as $m_2(y_1)$ for y_1 in $(k_1^S,\ k_2^S]$ and 0 otherwise. We, however, use separate notation, $n_2(y_1)$ and $m_2(y_1)$, to emphasize that we treat Stage II-N and Stage II-S separately when we design a procedure. The same argument is applied for treating w's and v's separately when they both represent the Stage II critical value functions.

Table 4.2 shows a flowchart of a two-stage adaptive procedure with a possible switch of the main study objective.

4.4 Determining Sample Sizes And Critical Values

We discuss how to determine the sample sizes and critical values for Stage I, Stage II-N and Stage II-S. For this purpose, the error probabilities discussed in Section 4.2 are further divided into Stage I

Table 4.2: *A flowchart of a general two-stage adaptive procedure with a possible switch of the main study objective.*



and Stage II error probabilities. For example, the error probability, $\alpha_{IN}(\xi_I)$ is defined in Section 4.2 as the probability of concluding N when the truth is $\xi = \xi_I$. The error, "IN" (Concluding N when the truth is I) can be made in Stage I, Stage II-N or Stage II-S. These error probabilities are denoted by $\alpha_{IN}^1(\xi_I)$, $\alpha_{IN}^{2N}(\xi_I)$ and $\alpha_{IN}^{2S}(\xi_I)$. The superscripts indicate the stage in which the error is made. And the sum of $\alpha_{IN}^{2N}(\xi_I)$ and $\alpha_{IN}^{2S}(\xi_I)$ is denoted by $\alpha_{IN}^{2}(\xi_I)$, so that $\alpha_{IN}(\xi_I) = \alpha_{IN}^1(\xi_I) + \alpha_{IN}^2(\xi_I)$.

Even though we use α's to represent these error probabilities in individual stages, we do not claim that these are the suprema of the error probabilities in their respective null spaces. For example, α_I^{2N} is not necessarily equal to $\sup_{\xi \le \xi_I} P_\xi$[Conclude N or S in Stage II-N]. We do not need P_ξ[Conclude N or S] or P_ξ[Conclude S] for each individual stage to be increasing in ξ in order to have the overall power functions $\psi_{I^c}(\xi)$ and $\psi_S(\xi)$ to be increasing in ξ.

4.4.1 Stage I

To specify a design of Stage I, we need to determine the sample size and critical values. The four critical values of Stage I satisfy the following:

$$\begin{aligned} \alpha^1_{IN}(\xi_I) &= P[\, k^I_2 < Y_1 \le k^S_1 \,|\, \xi = \xi_I \,] & \beta^1_N(\xi_N) &= P[\, Y_1 \le k^I_1 \,|\, \xi = \xi_N \,] \\ \alpha^1_{IS}(\xi_I) &= P[\, k^S_2 < Y_1 \,|\, \xi = \xi_I \,] & \beta^1_{SI}(\xi_S) &= P[\, Y_1 \le k^I_1 \,|\, \xi = \xi_S \,] \\ \alpha^1_N(0) &= P[\, k^S_2 < Y_1 \,|\, \xi = 0 \,] & \beta^1_{SN}(\xi_S) &= P[\, k^I_2 < Y_1 \le k^S_1 \,|\, \xi = \xi_S \,]. \end{aligned} \tag{4.4.7}$$

Using the fact that $Y_1 \sim \text{Normal}\,(\xi,\, 1/n_1)$, we can connect the critical values, sample size and error probabilities of Stage I[1]. These Stage I parameters cannot be specified independently of each other because they are connected through the relationships described in (4.4.7). For example, if n_1 and k^S_2 are given, then α^1_{IS} and α^1_N are both determined. More precisely, these components form the following three groups:

1. $\{n_1,\ k^S_2,\ \alpha^1_{IS}(\xi_I),\ \alpha^1_N(0)\}$,
2. $\{n_1,\ k^I_1,\ \beta^1_N(\xi_N),\ \beta^1_{SI}(\xi_S)\}$,
3. $\{n_1,\ k^I_2,\ k^S_1,\ \alpha^1_{IN}(\xi_I),\ \beta^1_{SN}(\xi_S)\}$.

Any two components from group 1, any two components from group 2 and any three components from group 3 are necesarry and sufficient to determine the remaining components within each respective group[2]. And these three groups are connected by the common component, n_1.

4.4.2 Stage II-N

For Stage II-N, define the "conditional power functions," $A^{I^c}(y_1, \xi)$ and $A^S(y_1, \xi)$, for $\xi \in (-\infty, \infty)$ as follows:

$$A^{I^c}(y_1, \xi) = P_\xi[\,\text{Conclude } N \text{ or } S \,|\, Y_1 = y_1], \tag{4.4.8}$$

$$A^S(y_1, \xi) = P_\xi[\,\text{Conclude } S \,|\, Y_1 = y_1]. \tag{4.4.9}$$

For all ξ and $y_1 \in (k^I_1, k^I_2\,]$, $A^{I^c}(y_1, \xi)$ and $A^S(y_1, \xi)$ take some value in $(0, 1)$. For $y_1 \in (-\infty, k^I_1\,]$, they are uniformly 0 and for $y_1 \in [\,k^I_2, \infty)$, they are uniformly 1. Because $A^{I^c}(y_1, \xi)$ and $A^S(y_1, \xi)$ are *conditional* power function, ξ is permitted to be a function, $\xi(y_1)$, of y_1. From the above definitions, for all $y_1 \in (k^I_1, k^I_2\,]$ we have $A^{I^c}(y_1, \xi) \ge A^S(y_1, \xi)$ with equality holding if and only if

[1] For example, $\alpha^1_{IS}(\xi_I) = 1 - \Phi[\sqrt{n_1}(k^S_2 - \xi_I)]$.

[2] In trivial situations such as when $\alpha^1_{IS}(\xi_I) = 0$, $\alpha^1_N(0) = 0$ and $k^S_2 = \infty$, the remaining component, n_1 in this case, cannot be determined.

$P_\xi[\,\text{Conclude } N \,|\, Y_1 = y_1\,] = 0$ at the particular ξ. We can express the Stage II-N error probabilities using A-functions. First, α's are:

$$\alpha_{IS}^{2N}(\xi_I) = \int_{k_1^I}^{k_2^I} A^S(y_1, \xi_I) g_{\xi_I}(y_1) dy_1, \tag{4.4.10}$$

$$\alpha_{I}^{2N}(\xi_I) = \int_{k_1^I}^{k_2^I} A^{I^c}(y_1, \xi_I) g_{\xi_I}(y_1) dy_1, \tag{4.4.11}$$

$$\alpha_{N}^{2N}(0) = \int_{k_1^I}^{k_2^I} A^S(y_1, 0) g_0(y_1) dy_1. \tag{4.4.12}$$

And β's are:

$$\beta_{N}^{2N}(\xi_N) = \int_{k_1^I}^{k_2^I} \left(1 - A^{I^c}(y_1, \xi_N)\right) g_{\xi_N}(y_1) dy_1, \tag{4.4.13}$$

$$\beta_{SI}^{2N}(\xi_S) = \int_{k_1^I}^{k_2^I} \left(1 - A^{I^c}(y_1, \xi_S)\right) g_{\xi_S}(y_1) dy_1, \tag{4.4.14}$$

$$\beta_{S}^{2N}(\xi_S) = \int_{k_1^I}^{k_2^I} \left(1 - A^{S}(y_1, \xi_S)\right) g_{\xi_S}(y_1) dy_1. \tag{4.4.15}$$

For these, we let $g_\xi(y_1)$ to be the pdf of Normal $(\xi,\, 1/n_1)$. We also have $\alpha_{IN}^{2N}(\xi_I)$ and $\beta_{SN}^{2N}(\xi_S)$, but we do not include these in the specification components. Each of these two error probabilities involves two conditional error functions, and they are not specified but rather calculated through $\alpha_{IN}^{2N}(\xi_I) = \alpha_{I}^{2N}(\xi_I) - \alpha_{IS}^{2N}(\xi_I)$, and $\beta_{SN}^{2N}(\xi_S) = \beta_{S}^{2N}(\xi_S) - \beta_{SI}^{2N}(\xi_S)$.

Using the decision rules for Stage II-N, (4.4.8) and (4.4.9) can be expressed as follows:

$$A^{I^c}(y_1, \xi) = P_\xi[\, Y_2 > w_1(y_1) \,|\, Y_1 = y_1], \tag{4.4.16}$$

$$A^{S}(y_1, \xi) = P_\xi[\, Y_2 > w_2(y_1) \,|\, Y_1 = y_1]. \tag{4.4.17}$$

Because the conditional distribution of Y_2 given $Y_1 = y_1$ is Normal $(\xi,\, 1/n_2(y_1))$, (4.4.16) and (4.4.17) can be further expressed in terms of $n_2(y_1)$, $w_1(y_1)$ and $w_2(y_1)$.

$$A^{I^c}(y_1, \xi) = 1 - \Phi\left[\sqrt{n_2(y_1)}\,(w_1(y_1) - \xi)\right], \tag{4.4.18}$$

$$A^{S}(y_1, \xi) = 1 - \Phi\left[\sqrt{n_2(y_1)}\,(w_2(y_1) - \xi)\right]. \tag{4.4.19}$$

Stage II-N is characterized by the following specification components: $n_2(y_1)$, $w_1(y_1)$, $w_2(y_1)$, $A^{I^c}(y_1, \xi)$ and $A^S(y_1, \xi)$, and these components are categorized into the following two groups:

1. $\{A^{I^c}(y_1, \xi_I), A^{I^c}(y_1, \xi^*), n_2(y_1), w_1(y_1)\ \}$
2. $\{A^{S}(y_1, \xi^\dagger), A^{S}(y_1, \xi^\ddagger), n_2(y_1), w_2(y_1)\ \}$

Note that the choice of ξ^*, which may be a function of y_1, is arbitrary as long as it is not equal to ξ_I. Also $\xi^\dagger$ and $\xi^\ddagger$ can be arbitrarily chosen as long as they are not equal to each other. The members of the first group are connected through (4.4.18), and the members of the second group are connected through (4.4.19), and specifying two components determines the remaining two components within each group[3]. These two groups are connected by the common component, $n_2(y_1)$.

For the first group we choose $A^{I^c}(y_1, \xi_I)$ to be one of the specification components as $A^{I^c}(y_1, \xi_I)$ is the most natural choice for a specification component because it gives, by (4.4.11), the probability, $\alpha_I^{2N}(\xi_I)$, which is necessary to control Type I error rate, $\alpha_I(\xi_I)$.

For the second group, unlike the first group, there is not a primary choice for $\xi^\dagger$ and for $\xi^\ddagger$. In order to control Type I error rate, it seems that $A^S(y_1, \xi_I)$ and $A^S(y_1, \xi_N)$ are reasonable choices. However, because $n_2(y_1)$ is usually determined by the first group, we only have one element to choose freely. Choosing an $A^S(y_1, \xi)$ at any ξ, or $w_2(y_1)$ in addition to $n_2(y_1)$ will determine the all the necessary elements.

4.4.3 Stage II-S

Stage II-S is analogous to Stage II-N. First we define, for Stage II-S, the "conditional power functions," $B^{I^c}(y_1, \xi)$ and $B^S(y_1, \xi)$, for $\xi \in (-\infty, \infty)$ as follows:

$$B^{I^c}(y_1, \xi) = P_\xi[\,\text{Conclude } N \text{ or } S \,|\, Y_1 = y_1],$$
$$B^S(y_1, \xi) = P_\xi[\,\text{Conclude } S \,|\, Y_1 = y_1].$$

The B-functions take values in $(0, 1)$ if $y_1 \in (k_1^S, k_2^S]$. If $y_1 \in (-\infty, k_1^S]$ then B-functions take the value of 0, and if $y_1 \in (k_2^S, \infty)$ then B-functions take the value of 1. The Stage II-S error probabilities can be expressed using the conditional power functions First, for α's,

$$\alpha_{IS}^{2S}(\xi_I) = \int_{k_1^S}^{k_2^S} B^S(y_1, \xi_I) g_{\xi_I}(y_1) dy_1, \tag{4.4.20}$$
$$\alpha_I^{2S}(\xi_I) = \int_{k_1^S}^{k_2^S} B^{I^c}(y_1, \xi_I) g_{\xi_I}(y_1) dy_1, \tag{4.4.21}$$
$$\alpha_N^{2S}(0) = \int_{k_1^S}^{k_2^S} B^S(y_1, 0) g_0(y_1) dy_1. \tag{4.4.22}$$

[3]Fact 2.2.2 is applicable with the four components within each group, and the actual algebraic connections within each group are similar to the ones described in equations (2.2.14) through (2.2.18).

And β's are:

$$\beta_N^{2S}(\xi_N) = \int_{k_1^S}^{k_2^S} \left(1 - B^{I^c}(y_1, \xi_N)\right) g_{\xi_N}(y_1) dy_1, \tag{4.4.23}$$

$$\beta_{SI}^{2S}(\xi_S) = \int_{k_1^S}^{k_2^S} \left(1 - B^{I^c}(y_1, \xi_S)\right) g_{\xi_S}(y_1) dy_1, \tag{4.4.24}$$

$$\beta_S^{2S}(\xi_S) = \int_{k_1^S}^{k_2^S} \left(1 - B^{S}(y_1, \xi_S)\right) g_{\xi_S}(y_1) dy_1. \tag{4.4.25}$$

Similarly to Stage II-N, we have the following two groups of specifying elements:

1. $\{B^{I^c}(y_1, \xi^\dagger), B^{I^c}(y_1, \xi^\ddagger), m_2(y_1), v_1(y_1)\}$,

2. $\{B^S(y_1, 0), B^S(y_1, \xi^*), m_2(y_1), v_2(y_1)\}$.

The primary objective for Stage II-N is to protect Type I error rate, $\alpha_{NS}^2(0)$, and in order to accomplish that objective, $B^S(y_1, 0)$ needs to be chosen so that it satisfies (4.4.22). Then one additional component from group 2, we can obtain the other two components in group 2. And with one additional component other than $m_2(y_1)$ specified for the group 1, Stage II-S is completely determined.

As long as the $A^{I^c}(y_1, \xi)$, $A^S(y_1, \xi)$, $B^{I^c}(y_1, \xi)$, $B^S(y_1, \xi)$ functions satisfy the conditions, (4.4.10) $\sim$ (4.4.15), and (4.4.20) $\sim$ (4.4.25), the probabilities of errors described in Section 4.2 are all controlled. Moreover, we can impose criteria such as maxima and minima on the Stage II sample sizes. With the additional restrictions on the Stage II sample sizes, not all of Type II error probabilities may be controlled. The criteria that must be satisfied are controlling $\alpha_I(\xi_I)$ and $\alpha_{NS}(0)$. Type II error probabilities, or the power, of the study may not be achievable with sample size restrictions. As discussed in Chapter 2, it is fairly easy to alter Stage II specification components so as to give rise to a particular design. And additionally, we can explore numerous designs and choose the one that is most satisfactory among them.

4.5 Literature Review

Two-stage adaptive procedures which test both noninferiority and superiority hypotheses are a new field of research, and there are only few papers that discuss such procedures. In this section, we consider papers by Wang et.al. (2001) and by Brannath et.al. (2003).

4.5.1 Wang et.al. (2001)

Wang et.al. (2001) consider adaptive multi-stage procedures, but they only allow one adaptation by which the main study objective and the sample size may be changed. Therefore, their procedure

is similar to the two-stage adaptive procedures that we consider in this dissertation. They have separate schemes for designing the study to show superiority as the main objective and noninferiority as the secondary objective (S-NI) and vice versa (NI-S). They let N and M be the group sequential procedure's sample sizes for the superiority objective and noninferiority objective, respectively.

They consider three types of procedures: group sequential stepwise test procedure (GSS), group sequential closed test procedure (GSC) and adaptive group sequential closed procedure (AGSC). GSS and GSC do not allow an interim sample size adjustment. Both GSS and GSC test for the main objective at each stage and if the superiority claim is reached, the study terminates. The difference between these two occurs when only the noninferiority conclusion is reached in an intermediate stage. In GSS, the procedure terminates with only the noninferiority claim, but in GSC, the procedure continues to the next stage without making any conclusion. Wang, et.al. (2001) cite the finding of Wang, et.al. (1999) that the unconditional power of GSS is very low for both superiority and noninferiority mainly because achievement of noninferiority will terminate the trial preventing testing further for superiority.

AGSC is a variation of GSC which allows for one interim adjustment (at an arbitrary time s) of the planned sample size based on the observed sample path. They only allow the sample size to *increase*, which indicates that the study's main objective for which the original sample size is determined needs to be the easier one of the two. In an S-NI situation, the decision to increase the total sample size from N to M is based on the conditional powers at the observed Δ. At the time s, they calculate $CP_S(\Delta)$ which is the conditional power for concluding superiority at the trial end with total sample size of N and $CP_E(\Delta)$ which is the conditional power for concluding noninferiority but not superiority at the trial end with total sample size of M. And if $CP_S(\Delta) < CP_E(\Delta)$, the sample size is increased from N to M, otherwise it remains N.

Wang, et.al. (2001) conduct simulation studies of the AGSC procedure with four possible interim analyses and one final analysis at equal time intervals before and after sample size adjustment. They set $\alpha = .025$ and $\beta = .20$ for both objectives and use $N = 330$ and $M = 1200$. The *a priori* chosen time of sample size adjustment s is varied among $\{.20, .40, .60, .80\}$. One alternative to *a priori* choosing s is that the conditional power criterion is checked at each interim analysis until the criterion is satisfied to increase the sample size at one interim analysis.

Wang, et.al. (2001) compute the Type I error rate (based on 50,000 simulations) and power (based on 10,000 simulations) for both superiority and noninferiority objectives. The Type I error rate ranges (for different choices of s) from .0250 to .0259 for the superiority objective and .0247 to .0255 for the noninferiority objective, and they claim that the Type I error rate of the AGSC test

is maintained at the targeted 0.025 level for superiority and noninferiority. The power ranges from .846 to .947 for the superiority objective and .607 to .768 for the noninferiority objective.

4.5.2 Brannath et.al. (2003)

Brannath et.al. (2003) consider the problem of simultaneous sequential tests for noninferiority and superiority. A unique feature of their procedure is that they consider a continuous hierarchical family of null hypotheses as follows: for $\Delta \in [0, \Delta_I]$,

$$H_0^{(\Delta)} : \mu_t - \mu_c = -\Delta,$$
$$H_1^{(\Delta)} : \mu_t - \mu_c > -\Delta,$$

where $\Delta_I > 0$ is the preassigned noninferiority margin. Possible conclusions of such a procedure are either inferiority (to accept all $H_0^{(\Delta)}$, Δ in $[0, \Delta_I]$), Δ^*-noninferiority (to reject all $H_0^{(\Delta)}$, Δ in $[\Delta^*, \Delta_I]$) or superiority (to reject all $H_0^{(\Delta)}$, Δ in $[0, \Delta_I]$).

Brannath et.al. (2003) consider a two-stage adaptive procedure in which the sample sizes for Stage I (n_1) and for Stage II (n_2) are fixed, but they claim that sample-size reassessments are possible within the framework they present. They also claim that in contrast to Wang et.al. (2001), Stage II (sample size) do not have to be changed to achieve sufficient overall power.

Their key result is proving that their procedure controls the multiple level α for the continuous family of null hypotheses $H_0^{(\Delta)}$, Δ in $[0, \Delta_I]$. In other words, this is the probability that at least one erroneous rejection within this family does not exceed α.

They treat the following two situations, where Δ_I is the noninferiority margin, and Δ_S is the point at which the superiority test is powered. In the first situation, $\Delta_S = 2\Delta_I$; here establishing noninferiority requires a much larger sample size than necessary for superiority. In the second situation, $\Delta_S = \Delta_I/2$, and the proof of superiority requires a larger sample size than necessary for noninferiority. They construct two designs for situation 1 and one design for situation 2 as examples. In each situation, $\alpha = .025$ and $\beta = .2$ for both superiority and noninferiority objectives. And N^S for situation (1) and N^I for situation (2) are both 1005 so that $\alpha = .025$ and $\beta = .20$. In the following paragraphs, we describe how to place their designs into our framework.

Example 1: Situation 1, "Ambitious" design: First the maximum total sample size is set at $N_I = 1005$. For Stage I, the probability of concluding S when $\Delta = \Delta_S$ is set to be .8. With the decision to use Pocock boundaries (Pocock 1977), $n_1 = 298$ and k_2^S are determined. From a condition to control the multiple level α, they obtain k_1^S. This procedure does not stop to conclude I and that there is only one Stage II (Stage II-S), therefore, $k_1^I = k_2^I = -\infty$. With this sample size and critical values, they calculate that the probability of concluding N and stop in Stage I when $\Delta = 0$ .014.

For Stage II-S, the sample size, n_2, is $N^I - n_1$ for all y_1. And the overall probability to conclude N or S when $\Delta = 0$ is set to be .8, and it is used to specify the Stage II-S probability to conclude N or S when $\Delta = 0$. Because the critical value $v_1(y_1)$ has a constant value for all y_1 in standardized (z) units, one can find $v_1(y_1)$ given n_2 and $P_{\Delta_0}[N \text{ or } S \text{ in Stage II } | Y_1 = y_1]$. The critical value function $v_2(y_1)$, which is constant for all y_1, is set to have the same standardized (z) value as k_2^S. Thus, the Stage II-S components are all specified.

Example 2: Situation 1, "Modest" design: Again the sample size limit is $N_I = 1005$. This design is similar to the "Ambitious" design above. The Stage I sample size remains at $n_1 = 298$. There is no Stage II-N again, so $k_1^I = k_2^I = -\infty$. They still specify the probability of concluding S when $\Delta = \Delta_S$ in Stage I to be .8 so that k_2^N is specified. The difference from the design of Example 1 is that in this design, there is higher probability of concluding N and terminating the study at the end of Stage I. By this change, the probability of continuing to Stage II is smaller than the "Ambitious" design, and the overall power to conclude N or S under $\Delta = 0$ decreases as a result. By adjusting the critical value for Stage I (k_1^S) and for Stage II-S ($v_1(y_1)$), we can make the overall power to conclude N or S under $\Delta = 0$ as high as .795. The other critical value function $v_2(y_1)$ is the same as for the "Ambitious" design.

Example 3: Situation 2, "Modest" design: The sample size for a conventional single-stage superiority procedure is $N^S = 1005$ when $\alpha = .025$ and power = .80. For this design, they again use the same $n_1 = 287$, and set the probability to conclude I^c (and stop) in Stage I to be .8. This determines k_1^S. In terms of the standardized values, k_1^S and k_2^S for this design are the same as k_1^S and k_2^S in Example 1. Also for this design, $k_1^I = k_2^I = -\infty$ and Stage II-N does not exist. The probability to conclude S when $\Delta = \Delta_S$ in Stage I is .019 by computation. For Stage II-S, $v_2(y_1)$ is chosen so that the overall power to conclude S when $\Delta = \Delta_S$ is .8. The other critical value $v_1(y_1)$ has the same standardized value as k_1^S for all y_1.

In examples 1 and 2, α, for both objectives, is controlled at .025 and the overall powers for the noninferiority hypothesis are .800 (Example 1) and .795 (Example 2). The overall powers for the superiority hypothesis are not controlled and they are virtually 1 in both examples. In Example 3, α for both objectives is controlled at .025, and the overall power for the superiority hypothesis is .800 and the overall power for the noninferiority hypothesis is virtually 1. Although constructing their procedure involves a complex numerical computation, the resulting designs fit in our framework.

4.6 Power Function Considerations

In this section, we consider the power functions of a two-stage adaptive procedure with possible switching of the main objectives. In Section 2.3, we establish the monotonicity of the power function of a two-stage adaptive procedure when there is no switching of the main objectives. The discussion in the current section is an extension of the discussion in Section 2.3.

In Section 4.2, we introduce two power functions: $\psi_{I^c}(\xi) = P_\xi[\text{conclude } N \text{ or } S]$ and $\psi_S(\xi) = P_\xi[\text{conclude } S]$. We need these functions to be bounded above by α for all $\xi \leq \xi_I$ and for all $\xi \leq 0$, respectively, to control the Type I error rates for the composite null hypotheses. An easy way to show that the power function is bounded above is to show that it is monotonically increasing in ξ. Monotonicity of the power function is also important in inference. Although not considered in this dissertation, it is our intention to develop methods to obtain unbiased estimators and confidence intervals for the difference of the means and the unconditional p-value of the test. We anticipate that the monotonicity of the power function will be a key property in developing the mentioned concepts for two-stage adaptive procedures.

In Chapter 2, we show that for a two-stage adaptive design with a single objective, the following two properties of the specification components are sufficient for the monotonicity of the power function, $\psi(\xi)$, in $-\infty < \xi \leq \xi^*$:

1. $A(y_1, \xi^*)$ is nondecreasing in $y_1 \in (k_1, k_2]$.

2. $n_2(y_1)$ is nonincreasing in $y_1 \in (k_1, k_2]$.

We first consider the monotonicity of $\psi_{I^c}(\xi)$. To replace the condition 1 above, we need that the conditional probability to conclude N or S is nondecreasing in y_1. What is not explicitly stated in the condition 1 above is that $A(y_1, \xi^*) = 0$ for y_1 in $(-\infty, k_1]$ and $A(y_1, \xi^*) = 1$ for y_1 in (k_2, ∞), making $A(y_1, \xi^*)$ nondecreasing in the entire range of y_1. We extend the conditional probability to conclude N or S to the entire range of y_1, and define $C^{I^c}(y_1, \xi)$ as follows:

$$C^{I^c}(y_1, \xi) = \begin{cases} 0 & \text{for } y_1 \text{ in } (-\infty,\ k_1^I] \\ A^{I^c}(y_1, \xi) & \text{for } y_1 \text{ in } (k_1^I,\ k_2^I] \\ 1 & \text{for } y_1 \text{ in } (k_2^I,\ k_1^S] \\ B^{I^c}(y_1, \xi) & \text{for } y_1 \text{ in } (k_1^S,\ k_2^S] \\ 1 & \text{for } y_1 \text{ in } (k_2^S,\ \infty). \end{cases}$$

An example of $C^{I^c}(y_1, \xi)$ is given in Figure 4.1. We observe that $C^{I^c}(y_1, \xi)$ is nondecreasing if $A^{I^c}(y_1, \xi)$ is nondecreasing in y_1 for all y_1 in $(k_1^I,\ k_2^I]$ and $B^{I^c}(y_1, \xi) = 1$ for all y_1 in $(k_1^S,\ k_2^S]$. If these two conditions and an additional condition that $n_2(y_1)$ is nonincreasing in $y_1 \in (k_1^I, k_2^I]$ are satisfied then we could use the analogous proofs to Lemma 2.3.1 and Lemma 2.3.2 to prove that

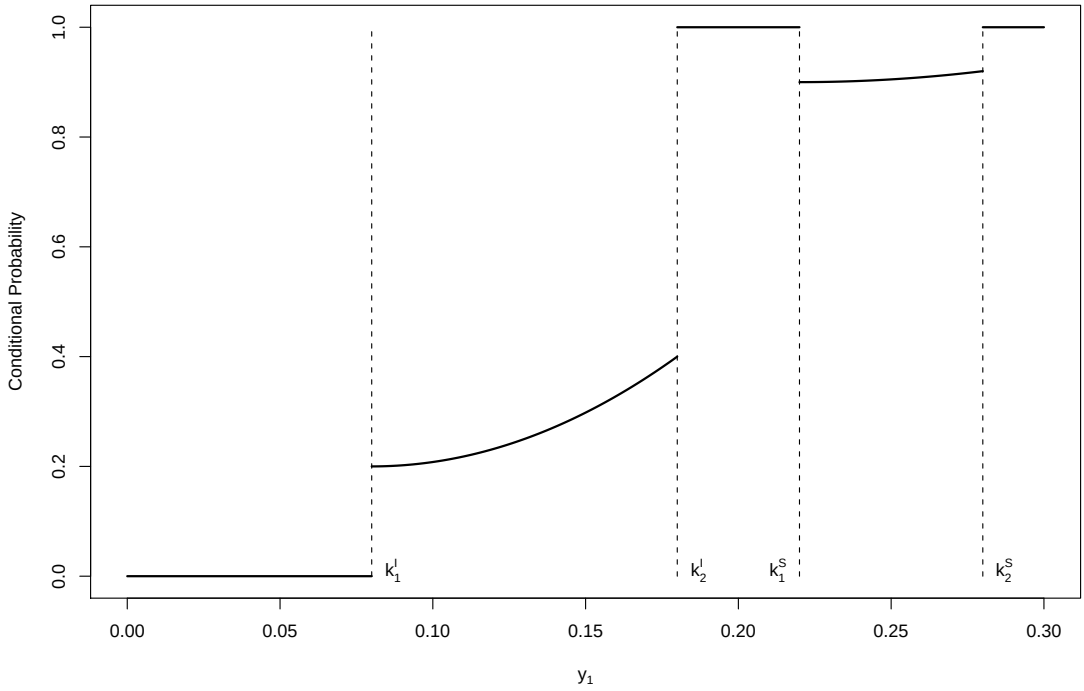


Figure 4.1: *An example of* $C^{I^c}(y_1, \xi)$.

- $\psi_{I^c}(\xi)$ is nondecreasing in ξ for all ξ in $(-\infty, \xi^*]$,
- $\psi_{I^c}(\xi) > \psi_{I^c}(\xi^*)$ for all ξ in $[\xi^*, \infty)$.

One of the conditions, $B^{I^c}(y_1, \xi) = 1$ for all y_1 in $(k_1^S, k_2^S]$, states that the probability to conclude N or S in Stage II-S is 1, which is accomplished by specifying $v_1(y_1) = -\infty$ for all y_1 in $(k_1^S, k_2^S]$. When $v_1(y_1) \neq -\infty$ we cannot use the analogous proofs to Lemma 2.3.1 and Lemma 2.3.2. These, however, are clearly not necessary conditions to provide monotonicity. The power function, $\psi_{I^c}(\xi)$, may possess the characteristics mentioned above even if $B^{I^c}(y_1, \xi) \neq 1$ or the other two conditions are not satisfied. So when these conditions are not satisfied, the monotonicity of the power function, $\psi_{I^c}(\xi)$, can be checked by computation.

The argument for the monotonicity of the other power function, $\psi_S(\xi)$ is analogous. We define $C^S(y_1, \xi)$ as the conditional probability to conclude S.

$$C^S(y_1, \xi) = \begin{cases} 0 & \text{for } y_1 \text{ in } (-\infty, k_1^I] \\ A^S(y_1, \xi) & \text{for } y_1 \text{ in } (k_1^I, k_2^I] \\ 0 & \text{for } y_1 \text{ in } (k_2^I, k_1^S] \\ B^S(y_1, \xi) & \text{for } y_1 \text{ in } (k_1^S, k_2^S] \\ 1 & \text{for } y_1 \text{ in } (k_2^S, \infty) \end{cases}$$

If $A^S(y_1, \xi^*) = 0$ for all y_1 in $(k_1^I, k_2^I]$, $B^S(y_1, \xi^*)$ is nondecreasing, then $C^S(y_1, \xi^*)$ is nondecreasing in y_1 for the entire range of y_1. With the condition that $m_2(y_1)$ is nonincreasing in y_1 for y_1 in $(k_1^S, k_2^S]$, we can apply the same argument as Lemmas 2.3.1 and 2.3.2 to conclude that

- $\psi_S(\xi)$ is nondecreasing in ξ for all ξ in $(-\infty, \xi^*]$,
- $\psi_S(\xi) > \psi_S(\xi^*)$ for all ξ in $[\xi^*, \infty)$.

However, even when $A^S(y_1, \xi^*) \neq 0$, we may still have monotonicity of $\psi_S(\xi)$, but we need to check it by computation.

4.7 Example

In the following example, we show how our framework is used to design a two-stage adaptive procedure with possible switching of the main objectives. The primary purpose of this hypothetical example is to illustrate our computational capability in designing a procedure which controls various error probabilities and the maximum sample sizes. This particular design's expected sample sizes are not satisfactory. We present in Section 4.8 another design which controls the error probabilities and possesses better characteristics in terms of expected sample sizes.

In this example, the primary objective is to show that a new treatment is noninferior to a competitor whose efficacy has already been established. The larger values of the observations indicate a preferable result. The noninferiority margin is $\Delta_I = 1$. The hypotheses to be tested are:

$$H_0^{NI} : \mu_t - \mu_c \leq -1 \qquad H_0^S : \mu_t - \mu_c \leq 0$$
$$H_1^{NI} : \mu_t - \mu_c > -1 \qquad H_1^S : \mu_t - \mu_c > 0$$

The power for the noninferiority objective is chosen to be calculated at $\Delta_N = 0$. The superiority margin is $\Delta_S = .5$. Because $\Delta_N = 0$, we do not have to make distinction between ξ_N and 0. Suppose that the population variances, σ_t^2 and σ_c^2, are both known to be 4^2; hence, $\sigma^2 = 4$. We standardize Δ's by the factor, $1/\sqrt{2}\,\sigma$ to get $\xi_I = -.1768$, $\xi_N = 0$ and $\xi_S = .0884$. Type I and Type II error probabilities for the procedure are set as follows: $\alpha_I = .025$, $\alpha_N = .025$, $\beta_N = .10$ and $\beta_S = .20$. With these error probabilities and Δ's specified, we can calculate N_I and N_S, the single-stage procedure's sample sizes for noninferiority and superiority objectives, respectively. They are $N_I = 337$ and $N_S = 1005$.

We choose to take 80% of N_I in Stage I, so that $n_1 = .8 \times 337 = 270$. Furthermore, we choose the Stage I critical values that satisfy the following:

$$P[Y_1 < k_1^I \,|\, \xi = \xi_N] = P[\text{Conclude } I \,|\, \xi = \xi_N] = .025,$$
$$P[k_2^I < Y_1 \,|\, \xi = \xi_I] = P[\text{Conclude } N \text{ or } S \text{ or continue to Stage II-S} \,|\, \xi = \xi_I] = .010,$$
$$P[Y_1 < k_1^S \,|\, \xi = \xi_S] = P[\text{Conclude } I \text{ or } N \text{ or continue to Stage II-N} \,|\, \xi = \xi_S] = .050,$$
$$P[k_2^S < Y_1 \,|\, \xi = \xi_N] = P[\text{Conclude } S \,|\, \xi = \xi_N] = .005.$$

Then we have $k_1^I = -.1193$, $k_2^I = -.0352$, $k_1^S = -.0117$ and $k_2^S = .1568$. For the groups noted in Section 4.4.1, we have two components from group 1 and from group 2, and three components from group 3. Therefore, we can obtain all the necessary Stage I components using (4.4.7) as listed below:

1. $\{n_1 = 270,\ k_2^S = .1568,\ \alpha_{IS}^1(\xi_I) = .000,\ \alpha_N^1(0) = .005\}$,
2. $\{n_1 = 270,\ k_1^I = -.1193,\ \beta_N^1(\xi_N) = .025,\ \beta_{SI}^1(\xi_S) = .000\}$,
3. $\{n_1 = 270,\ k_2^I = -.0352,\ k_1^S = -.0117,\ \alpha_{IN}^1(\xi_I) = .007,\ \beta_{SN}^1(\xi_S) = .029\}$.

These components specify a unique Stage I. The following Table 4.3 summarizes the Stage I probabilities.

Table 4.3: *Example of Section 4.7: Stage I probabilities.*

ξ	Conclude I and stop	Continue to Stage II-N	Conclude N and stop	Continue to Stage II-S	Conclude S and stop
ξ_I	.828	.162	.007	.003	.000
ξ_N	.025	.257	.142	.571	.005
ξ_S	.000	.021	.029	.819	.131

$\xi_I = -.1768$, $\xi_N = 0$ *and* $\xi_S = .0884$.

Now we proceed to planning Stage II-N and Stage II-S. Before specifying the components for Stage II-N and for Stage II-S, we need to set criteria for the error probabilities for Stage II's. Some of the Stage II probabilities can be specified, and others are obtained through simple calculation[4] to satisfy the overall error probability conditions such as $\alpha_I(\xi_I) = .025$. In Table 4.4, the Stage II probabilities that are specified are in the bold font, and those that result through calculation are in italic font. Probabilities when $\xi = \xi_S$ depend on the sample size functions and critical value functions for Stage II-N and Stage II-S, and they cannot yet be determined.

[4]For example, we choose for Stage II-N, $P[\text{Conclude } I \text{ in Stage II-N} \,|\, \xi = \xi_N] = .050$ and $P[\text{Conclude } S \text{ in Stage II-N} \,|\, \xi = \xi_N] = .007$. Then $P[\text{Conclude } N \text{ in Stage II-N} \,|\, \xi = \xi_N] = .257 - .050 - .007 = .200$. Furthermore, because $\alpha_N(\xi_N) = .025$, we need to have $\alpha_N^{2S}(\xi_N) = \alpha_N(\xi_N) - \alpha_N^1(\xi_N) - \alpha_N^{2N}(\xi_N) = .025 - .005 - .007 = .013$.

Table 4.4: *Example of Section 4.7: Unconditional Stage I and Stage II probabilities.*

Stage I	I	II-N	Stage II-N			N	II-S	Stage II-S			S
Stage II	-	-	I	N	S	-	-	I	N	S	-
$\xi = \xi_I$	.828	.162	*.144*	*.018*	**.000**	.007	.003	**.003**	*.000*	**.000**	.000
$\xi = \xi_N$	.025	.257	**.050**	*.200*	**.007**	.142	.571	*.025*	*.533*	*.013*	.005
$\xi = \xi_S$	.000	.021				.029	.819				.131

$\xi_I = -.1768$, $\xi_N = 0$ *and* $\xi_S = .0884$.

I, N, S are the three conclusions.

II-N and II-S are the Stage I probabilities of continuing to the respective Stage II.

For Stage II-N, we impose sample size restrictions so that the total sample size, $n_1 + n_2(y_1)$ is between $N^I = 337$ and $1.25 \times N^I = 421$. We choose[5] the following A^{I^c}-functions, which are displayed in Figure 4.2:

$$A^{I^c}(y_1, \xi_I) = .04 + 57.05(y_1 - k_1^I)^2,$$

$$A^{I^c}(y_1, \xi_N) = \begin{cases} 1 - \Phi\left[z_{A^{I^c}(y_1,\xi_I)} + \sqrt{151}\,(\xi_I - \xi_N)\right] & \text{if } k_1^I < y_1 < -.100 \\ .70 + 1.917(y_1 - k_1^I) & \text{if } -.100 \le y_1 < -.0496 \\ 1 - \Phi\left[z_{A^{I^c}(y_1,\xi_I)} + \sqrt{67}\,(\xi_I - \xi_N)\right] & \text{if } -.0496 \le y_1 < k_2^I. \end{cases}$$

With these two A^{I^c} functions, we can obtain $n_2(y_1)$, $w_1(y_1)$ and $A^{I^c}(y_1, \xi)$ for any ξ using the functional connections among the first group: $\{A^{I^c}(y_1, \xi_I),\ A^{I^c}(y_1, \xi^*),\ n_2(y_1),\ w_1(y_1)\}$. The total sample size, $n_1 + n_2(y_1)$ is displayed in Figure 4.3.

For the second group, we only have one free component to choose because $n_2(y_1)$ is chosen by the first group. We choose

$$A^S(y_1, \xi_N) = .001 + .5232(y_1 - k_1^I),$$

which is a straight line from .001 at k_1^I to .045 at k_2^I. All the components, $\{A^S(y_1, \xi^\dagger),\ A^S(y_1, \xi^\ddagger),\ n_2(y_1),\ w_1(y_1)\}$ of the second group are determined. Then using $A^S(y_1, \xi)$ for $\xi = \xi_I$, ξ_N and ξ_S, we can calculate P[Conclude S in Stage II-N] at these ξ's. It is .000 when $\xi = \xi_I$, .007 when $\xi = \xi_N$ and .003 when $\xi = \xi_S$. The Stage II-N probabilities when $\xi = \xi_S$ are now completely determined. See Table 4.5.

Now we proceed to designing Stage II-S. Because $1 - \beta_S = .8$ and P[Conclude S in Stage I $|\, \xi = \xi_S$] = .131 and P[Conclude S in Stage II-N $|\, \xi = \xi_S$] = .003, we need to have P[Conclude S in Stage II-S $|\, \xi = \xi_S$] = $.8 - .131 - .003 = .666$. For Stage II-S, first we impose the sample size restriction that the total sample size is between $N^S = 1005$ and $1.5 \times N^S = 1508$. We choose the

[5]First we chose $A^{I^c}(y_1, \xi_I) = .04 + 57.05(y_1 - k_1^I)^2$ and $A^{I^c}(y_1, \xi_N) = .70 + 1.917(y_1 - k_1^I)$. Then $A^{I^c}(y_1, \xi_N)$ is modified to accommodate the sample size restrictions. See Section 2.4.2 for a detailed description of the technique.

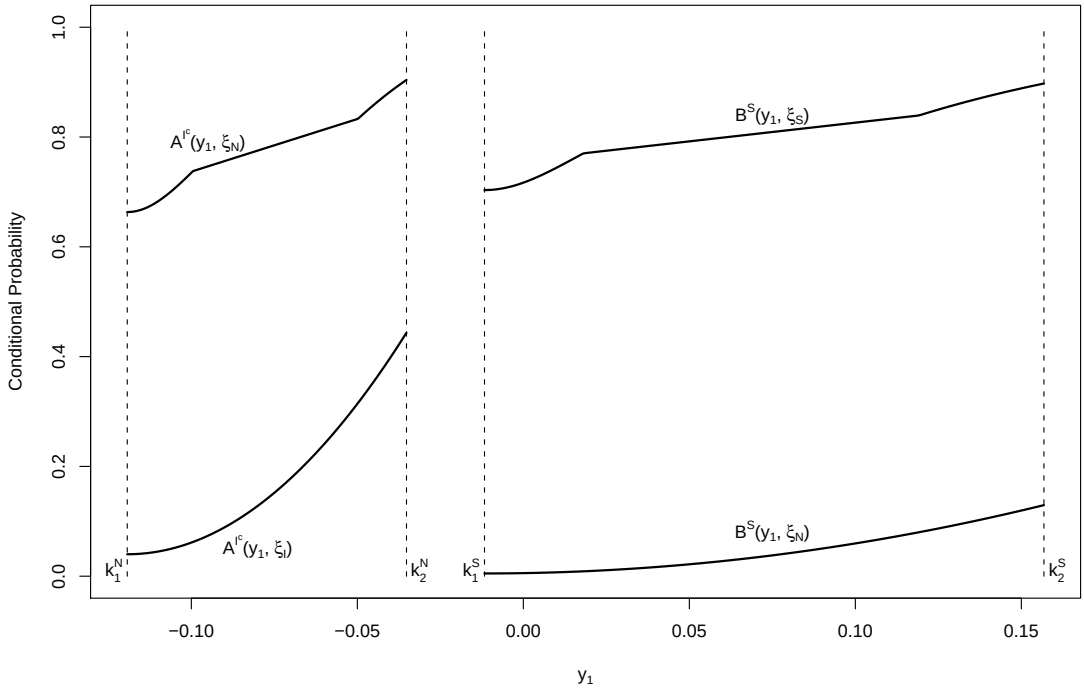


Figure 4.2: *Example of Section 4.7:* $A^{I^c}(y_1, \xi)$ *and* $B^S(y_1, \xi)$.

following:

$$B^S(y_1, 0) = .005 + 4.383(y_1 - k_1^S)^2,$$
$$B^S(y_1, \xi_S) = \begin{cases} 1 - \Phi\left[z_{B^S(y_1,\xi_N)} + \sqrt{1238}\,(\xi_N - \xi_S)\right] & \text{if } k_1^S < y_1 < .0171 \\ .75 + .6647(y_1 - k_1^S) & \text{if } .0171 \le y_1 < .1179 \\ 1 - \Phi\left[z_{B^S(y_1,\xi_N)} + \sqrt{835}\,(\xi_N - \xi_S)\right] & \text{if } .1179 \le y_1 < k_2^S. \end{cases} \tag{4.7.26}$$

See Figures 4.2 and 4.3 for the B functions and $m_2(y_1)$. The second group $\{B^S(y_1, 0),\ B^S(y_1, \xi^*),$

Table 4.5: *Example of Section 4.7: The Complete probabilities.*

Stage I	I	II-N	Stage II-N			N	II-S	Stage II-S			S
Stage II	-	-	I	N	S	-	-	I	N	S	-
$\xi = \xi_I$	.828	.162	.144	.018	.000	.007	.003	.003	.000	.000	.000
$\xi = \xi_N$	.025	.257	.050	.200	.007	.142	.571	.025	.533	.013	.005
$\xi = \xi_S$	.000	.021	.001	.017	.003	.029	.819	.000	.153	.666	.131

$\xi_I = -.1768$, $\xi_N = 0$ *and* $\xi_S = .0884$.

$m_2(y_1),\ v_2(y_1)\}$ is completely determined. For the first group, we specify

$$B^{I^c}(y_1, \xi_N) = .94 + .5945(y_1 - k_1^S),$$

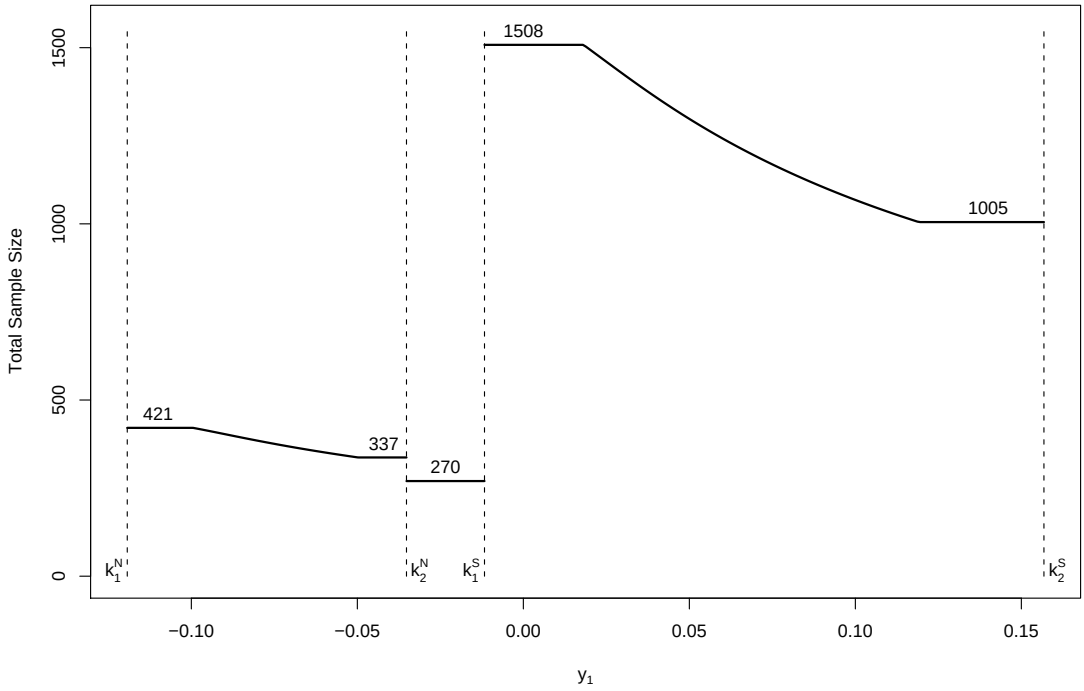


Figure 4.3: *Example of Section 4.7:* $n_1 + n_2(y_1)$ *and* $n_1 + m_2(y_1)$.

which is a straight line function from .94 at $y_1 = k_1^S$ to .99 at $y_1 = k_2^S$. With this $B^{I^c}(y_1, \xi_N)$ and $m_2(y_1)$ from the second group, all the components of the first group are determined. Then we can calculate all the Stage II-S probabilities in Table 4.5 when $\xi = \xi_S$. The critical value functions for Stage II-N and for Stage II-S are displayed in Figure 4.4. With the exception of $w_2(y_1)$, all are decreasing in y_1, which makes practical sense because within each continuation region, a larger y_1 suggests a stronger evidence against H_0^{NI} in favor of H_1^{NI} and against H_0^S in favor of H_1^S.

This particular design gives the following characteristics:
$\alpha_I(\xi_I) = \alpha_N(0) = .025$, $\alpha_{IN}(\xi_I) = .025$ and $\alpha_{IS}(\xi_I) = .000$;
$\beta_N(\xi_N) = .1$, $\beta_S(\xi_S) = .2$, $\beta_{SN}(\xi_S) = .199$ and $\beta_{SI}(\xi_S) = .001$.

Finally, the expected sample sizes are tabulated in Table 4.6. Let

$$E^*[n_2(y_1)] = E[n_2(y_1) \,|\, \text{Continue to Stage II-N}],$$
$$E^*[m_2(y_1)] = E[m_2(y_1) \,|\, \text{Continue to Stage II-S}].$$

The probabilities of continuing and the conditional expected sample sizes, $E^*[n_2(y_1)]$ and $E^*[m_2(y_1)]$ are tabulated for Stage II-N and Stage II-S. The *unconditional* total sample sizes in the last column

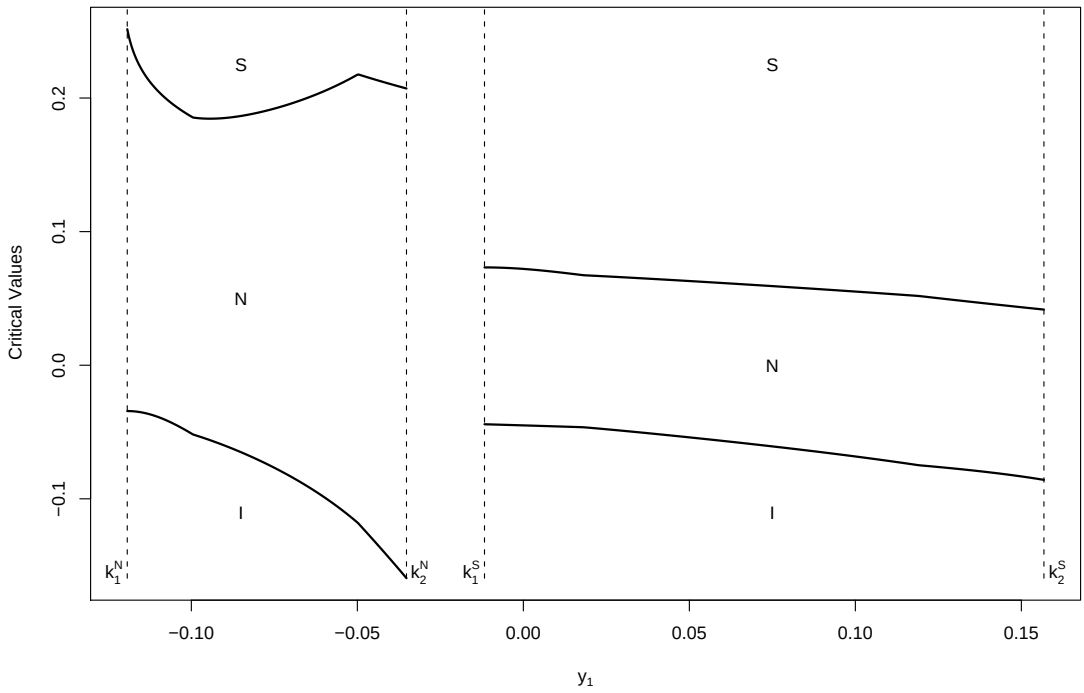


Figure 4.4: *Example of Section 4.7: Critical value functions.*

are calculated as follows:

$$E[\text{Total Sample Size}] = n_1 + E^*[n_2(y_1)] \times P[\text{Continue to Stage II-N}] + E^*[m_2(y_1)] \times P[\text{Continue to Stage II-S}].$$

Table 4.6: *Example of Section 4.7: Expected Sample Sizes.*

ξ	n_1	P[II-N]	$E^*[n_2(y_1)]$	P[II-S]	$E^*[m_2(y_1)]$	E[Sample Size]
$\xi = \xi_I$	270	.162	129.0	.003	1366.7	295.0
$\xi = \xi_N$	270	.257	95.3	.571	1080.2	911.3
$\xi = \xi_S$	270	.021	81.0	.819	911.7	1018.4

$\xi_I = -.1768$, $\xi_N = 0$ *and* $\xi_S = .0884$.

In the above example, our framework is able to build a design which controls the error probabilities when $\xi = \xi_I$, ξ_N, ξ_S and the maximum total sample sizes. In this design, both P[Conclude S in Stage II-N] and P[Conclude I in Stage II-S] are not equal to 0. Therefore, we cannot apply the argument in Section 4.6. However, by numerical computation, we find that both $\psi_{I^c}(\xi)$ and $\psi_S(\xi)$ are nondecreasing in the entire range of ξ.

4.8 A Variation Of The Design

Various criteria may be used in choosing a design, and small expected sample sizes are an important criterion. In the example in Section 4.7, the expected sample sizes are 295.0, 911.3, 1018.4 when $\xi = \xi_I$, ξ_N, ξ_S, respectively. The single-stage conventional design's sample sizes are $N_I = 327$ and $N_S = 1005$. When $\xi = \xi_I$, the expected sample size's percentage reduction from N_I is about 10 %. When $\xi = \xi_S$, the expected sample size is actually bigger than N_S by about 1.7 %. In Chapter 2, it is shown by a computational experiment that a small n_1 relative to N yields unfavorable results in terms of expected sample size. In this example, for the superiority objectives, $n_1 = 270$ is only 26.8 % of N_S. It can be argued that because we test for both noninferiority and superiority, we should compare the sample size of the two-stage adaptive procedure to $N_I + N_S = 1342$. Then the percentage reduction in terms of expected sample size when $\xi = \xi_S$ is $(1342 - 1018)/1342 \approx 25\%$.

When $\xi = \xi_N$, the expected sample size is 911.3, which is about 2.8 times as large as N_I. One reason for the large expected sample size is that we move on to Stage II-S with a fairly high probability of .571 when $\xi = \xi_N$. In Stage II-S the sample size, $m_2(y_1)$, is at least 735. In order to make the expected total sample size smaller, therefore, we propose the following two changes. 1) make n_1 larger, and 2) make $P[\text{Continue to Stage II-S} \,|\, \xi = \xi_N]$ smaller.

We consider the same setting as the last example and present a design that has larger n_1 and smaller $P[\text{Continue to Stage II-S} \,|\, \xi = \xi_N]$. We choose $n_1 = 337$ which is the same as N_I. Then because there should be enough information for the noninferiority objective in Stage I, we chose not to have Stage II-N. Setting $P[\, Y_1 > k_2^I \,|\, \xi = \xi_I \,] = .025$, we calculate $k_2^I = -.0699$. Because the sample size, N_I is chosen so that $\alpha = .025$ and $\beta = .10$ for the noninferiority objective, $\beta_N^1 = .10$ when $k_1^I = k_2^I = -.0699$. The other two critical values for Stage I, k_1^S and k_2^S are chosen so that $P[\, Y_1 < k_1^S \,|\, \xi = \xi_S \,] = .15$ and $P[\, k_2^S < Y_1 \,|\, \xi = \xi_N \,] = .010$. In the last example of Section 4.7, these probabilities are .05 and .005, respectively. Both of these changes make the continuation region narrower. We have $k_1^S = .0319$ and $k_2^S = .1267$. The Stage I probabilities are now determined, and they are tabulated in Table 4.7. $P[\text{Continue to Stage II-S} \,|\, \xi = \xi_N]$ is .571 under the original design,

Table 4.7: *Example of Section 4.8: Stage I probabilities*

ξ	Conclude I and stop	Continue to Stage II-N	Conclude N and stop	Continue to Stage II-S	Conclude S and stop
ξ_I	.975	0	.025	.000	.000
ξ_N	.100	0	.621	.269	.010
ξ_S	.002	0	.148	.609	.241

$\xi_I = -.1768$, $\xi_N = 0$ *and* $\xi_S = .0884$.

but it is only .269 with the new design.

For Stage II-S, we impose the same sample size constraint as in Section 4.7, that is, the total sample size is between $N^S = 1005$ and $1.5 \times N^S = 1508$. We also note in Table 4.7 that $P[\text{Continue to Stage II-S} \,|\, \xi = \xi_I] = 0$. Therefore, there is no need to try to conclude I in Stage II-S. So we let $v_1(y_1) = -\infty$ for all y_1 in the continuation region. We choose the following $B^S(y_1, \xi_N)$ and $B^S(y_1, \xi_S)$:

$$B^S(y_1, \xi_N) = .01 + 31.03(y_1 - k_1^S)^2,$$

$$B^S(y_1, \xi_S) = \begin{cases} 1 - \Phi\left[z_{B^S(y_1,\xi_N)} + \sqrt{1171}\,(\xi_N - \xi_S)\right] & \text{if } k_1^S < y_1 < .0757 \\ .90 + .90(y_1 - k_1^S) & \text{if } .0757 \le y_1 < .118 \\ 1 - \Phi\left[z_{B^S(y_1,\xi_N)} + \sqrt{941}\,(\xi_N - \xi_S)\right] & \text{if } .118 \le y_1 < k_2^S. \end{cases}$$

These B^S functions are displayed in Figure 4.5. We cannot use the same B^S functions as (4.7.26)

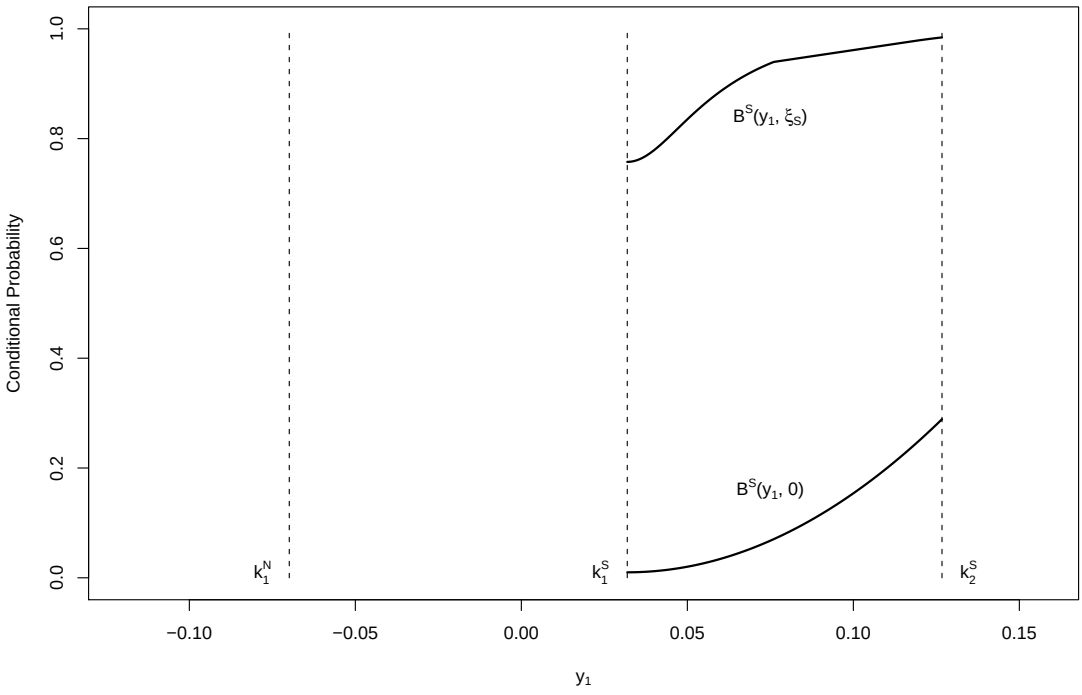


Figure 4.5: *Example of Section 4.8:* $B^S(y_1, \xi)$.

because the continuation region is different. However, in light of the result of the ANOVA study in Section 2.5, we conclude that using different B^S functions do not greatly affect the characteristics (i.e., the powers at various ξ's and expected sample size) of the resulting designs. The sample size functions are displayed in Figure 4.6. And the final probabilities for this design are tabulated in Table 4.8, and the expected sample sizes are tabulated in Table 4.9. This design's error probabilities are:

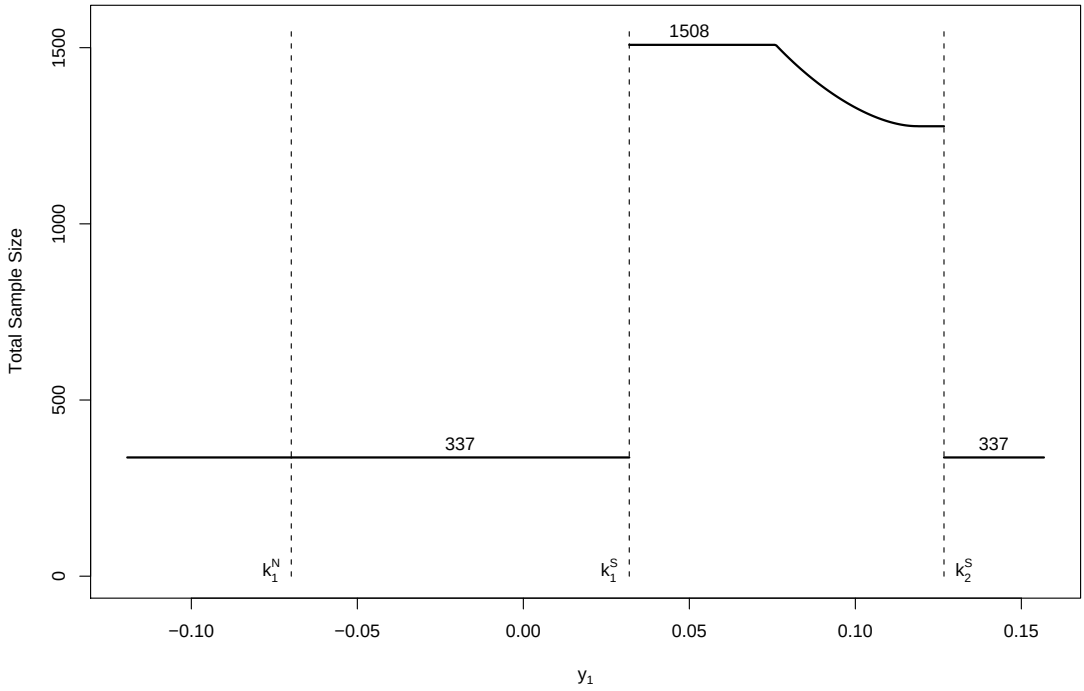


Figure 4.6: *Example of Section 4.8:* $n_1 + m_2(y_1)$.

Table 4.8: *Example of Section 4.8: Unconditional Stage I and Stage II probabilities*

Stage I	I	II-N	Stage II-N			N	II-S	Stage II-S			S
Stage II	-	-	I	N	S	-	-	I	N	S	-
ξ_I	.975	0	-	-	-	.025	.000	0	.000	.000	.000
ξ_N	.100	0	-	-	-	.621	.269	0	.254	.015	.010
ξ_S	.002	0	-	-	-	.148	.609	0	.050	.559	.241

$\xi_I = -.1768$, $\xi_N = 0$ *and* $\xi_S = .0884$.

$\alpha_I(\xi_I) = \alpha_N(0) = .025$, $\alpha_{IN}(\xi_I) = .025$ and $\alpha_{IS}(\xi_I) = .000$;

$\beta_N(\xi_N) = .1$, $\beta_S(\xi_S) = .2$, $\beta_{SN}(\xi_S) = .198$ and $\beta_{SI}(\xi_S) = .002$.

The maximum sample size is controlled at the same limit (1508) as the original design and expected sample sizes are more preferable in this design than in the original design. Especially, when $\xi = \xi_N$, the expected sample size is reduced from 911.3 to 642.9.

The power functions are easier to analyze in this example than the one in Section 4.7. Because there is no Stage II-N, by the facts that $m_2(y_1)$ is nonincreasing and $B^S(y_1, \xi_S)$ is nondecreasing, we have, by Lemma 2.3.2, that

1. $\psi_S(\xi)$ is nondecreasing for ξ in $(-\infty, \xi_S]$.

Table 4.9: *Example of Section 4.8: Expected Sample Sizes.*

ξ	n_1	P[II-N]	$E^*[n_2(y_1)]$	P[II-S]	$E^*[m_2(y_1)]$	E[Sample Size]
$\xi = \xi_I$	337	-	-	.000	0.0	337.0
$\xi = \xi_N$	337	-	-	.269	1137.2	642.9
$\xi = \xi_S$	337	-	-	.609	1081.1	995.4

$\xi_I = -.1768$, $\xi_N = 0$ *and* $\xi_S = .0884$.

2. $\psi_S(\xi) \geq \psi_S(\xi_S)$ for ξ in (ξ_S, ∞).

Moreover, by numerical computation, we find that $\psi_S(\xi)$ is also nondecreasing for ξ in (ξ_S, ∞). For $\psi_{I^c}(\xi)$, the argument is simpler. There is no Stage II-N, and we do not conclude I in Stage II-S. Therefore, $\psi_{I^c}(\xi)$ is equal to $P_\xi[k_1^S < y_1]$. Because the distribution of Y_1 is Normal $(\xi, 1/n_1)$, $\psi_{I^c}(\xi)$ is increasing in ξ.

4.9 Discussion

In this chapter, we consider two-stage adaptive procedures which test both noninferiority and superiority hypotheses. In addition to proper control of Type I error rates, the experimenter can specify the power for both objectives using our framework. Although testing noninferiority and superiority simultaneously in one trial (either adaptive or non-adaptive procedures) has been the topic of several papers, procedures which allow specification of the power for both objectives are rarely discussed.

In this setting in which these two study objectives are considered simultaneously, because there are two sets of hypotheses, we must define two separate power functions, one for the noninferiority objective and the other for the superiority objective. With the two power functions, we can specify Type I error rate and Type II error rate for each of the two sets of hypotheses. We use these error rates, as well as the expected and maximum sample sizes, as criteria to judge various designs.

A two-stage adaptive procedure with a possible switching of the main study objectives is more complicated than the basic procedure we consider in Chapter 2. The main purpose of the framework presented in this chapter is to facilitate such a design. Using our framework, one can construct a design that controls probabilities of Type I error and Type II error for each of the noninferiority and superiority objectives.

Our framework offers flexibility in designing a two-stage adaptive procedure. It may be desirable to design a procedure that does not stop for futility (to conclude inferiority) in Stage I, or a procedure

that does not conclude inferiority once it moves on to Stage II-S with superiority as the main study objective. Using our framework, it is possible to design a procedure with these design specifications.

We presented two designs as examples in this chapter. The first example in Section 4.7 is an example of the most general type of design in which all the decisions (I , N or S) are possible at the end of each of Stages I, II-N and II-S. This design is fairly complex because there are four critical values in Stage I, and two critical values in each of the Stage II's. This example illustrates how to construct such a design which controls Type I and Type II errors for both of the noninferiority and superiority objectives.

We note that for the example in Section 4.7, although we control various error rates, the expected sample sizes may not be satisfactory. Specifically, when $\xi = \xi_N$, the expected sample size is about 2.8 times as large as the sample size of a single-stage procedure with the main study objective of showing noninferiority. Therefore, with the aim of reducing the expected sample size when $\xi = \xi_N$, we construct the second design in Section 4.8. This design is simpler than the first one because Stage II-N does not exist, and concluding I in Stage II-S is not possible. The error rates of this design are compatible to those of the first, and the expected sample sizes when $\xi = \xi_N$ and when $\xi = \xi_S$ are smaller. Thus, we conclude that the second design is preferable to the first one.

When evaluating a two-stage adaptive design's expected sample size, comparing it to a single-stage procedure's sample size, N_I or N_S, may not be a fair comparison. To illustrate this idea, consider an extreme case in which $\Delta_S - 0$ is only one tenth of $\Delta_N - \Delta_I$. Then, if α and β are the same for both objectives, N_S would be 100 times as large as N_I. The Stage II-S sample size for a two-stage adaptive design would also be very large compared to N_I. Then even with a very small probability of continuing to Stage II-S, the expected sample size of the two-stage adaptive design would be much larger than N_I.

We need to design a two-stage adaptive procedure carefully to avoid some undesirable features in terms of error rates and expected sample sizes. We do not attempt to answer the question of optimality because the desired goals may vary from case to case. Our framework offers a tool by which different designs with various specific goals can be constructed.

Chapter 5

Future Research Topics

5.1 The Basic Framework

In this dissertation, we present a framework for two-stage adaptive procedures. The basic framework is established in Chapter 2. There are two topics in Chapter 2 that we would like to explore more in the future. The first of these is the power function of two-stage adaptive procedures. We give sufficient conditions on Stage II specification components which guarantee that the power function of the design is nondecreasing up to some point and bounded below beyond that point. As noted in Chapter 2, these conditions are not necessary, and many procedures that do not satisfy these two conditions still yield a nondecreasing power function. The example in Section 2.3 gives a procedure with a power function which is not nondecreasing. This particular design is extreme in the sense that the conditional power when y_1 is much smaller than ξ_0 is .90 under the null hypothesis. Therefore, even the Stage I data favor the null over the alternative; the conditional probability to reject H_0 in Stage II is very high. This example shows that the power function can be nondecreasing in ξ for a two-stage adaptive design, but it also implies that a design needs to be extreme like this one to make the power function decreasing in ξ in some interval. With examples like this in mind we would like to find necessary conditions that guarantee the power function is nondecreasing in ξ.

The second topic that we hope to revisit is the study concerning the design parameters of Section 2.5. The question of the optimal design has not been answered, and it is probably difficult to answer it analytically. Therefore, a study like the one in Section 2.5 is important to shed light on the question of optimality. Although the study in this dissertation provides the important findings that the shape of the conditional power function has little effect on the design characteristics, we would like to continue and perform a more thorough and systematic experiment. For example, in this study, we only considered $\alpha_1 = .01$, .025 and .04. So all of the procedures in this study stop at the end of Stage I to reject H_0. Some of the procedures in the literature [e.g., Lan and Trost (1997)] have $\alpha_1 = 0$, i.e., the procedure does not stop at the end of Stage I to reject H_0 at all. We would

like to study the effect of letting $\alpha_1 = 0$. Similarly, a procedure which does not stop at the end of Stage I for futility is equally of interest. Some papers [e.g., Gould (2001)] distinguish procedures with and without early stopping at the end of Stage I. From our perspective, all these procedures can be placed in one formulation and studied together.

5.2 Practical Considerations

Chapter 3 deals with practical issues of unequal sample sizes (including missing observations) and unknown variances. The solutions to these problems that we present in this dissertation are clearly not the only ones. The treatment of unequal sample sizes that we provide is satisfactory, but there may be better techniques to solve this problem. The unknown variances problem is divided into two situations: large Stage I sample sizes and small Stage I sample sizes. The former situation is treated in the same manner as missing observations are treated, and it is a reasonable procedure. However, we did not fully investigate the latter situation, i.e., when the true variances are unknown and Stage I sample sizes are small. We present an approach in Section 3.3.5, but there are issues that we have not considered. For example, at the end of Stage I, we have an estimate, s_{1p}^2, of the variance. We also have a prestudy estimate, σ_0^2. Because the Stage I sample size is small, the investigator could decide to use σ_0^2 by totally abandoning s_{1p}^2. Or the upper bound of a confidence interval for σ^2 using Stage I data may be used as a pessimistic estimate of σ_0^2. The decision as to which of these estimators to use may depend on the investigator's prior confidence concerning σ_0^2.

5.3 Switching The Main Study Objectives

Chapter 4 presents an extension of the basic framework to a practically motivated problem of testing noninferiority and superiority in the same trial. We present a framework that handles this situation, and the framework achieves our primary goal of being able to facilitate a design. Using the framework, one can design a fairly complicated design with the two study objectives. Design of a two-stage adaptive procedure that allows switching the main study objective is more complicated than the original framework for one study objective. Our framework is constructed to give a very general design. A topic for future research is to study some special designs. The following is the list of the special designs we would like to consider. First regarding the Stage I design:

- $k_1^I = -\infty$: The procedure does not stop at the end of Stage I for futility, i.e., to conclude I.
- $k_2^S = \infty$: To stop and conclude superiority, S, at the end of Stage I is not possible.

- $k_2^I = k_1^S$: The procedure does not stop at the end of Stage I to conclude "noninferiority but not superiority," N. It always moves on to Stage II unless y_1 is extreme in either direction, concluding inferiority, I, or S in Stage I.
- $k_1^I = k_2^I$: Stage II-N does not exist.
- $k_1^S = k_2^S$: Stage II-S does not exist.

We may also use some combinations of these. The sample size n_1 is a part of the Stage I design. So our investigation whould include the choice of n_1 also. Some special cases for Stage II are:

- $w_2(y_1) = \infty$: In Stage II-N, concluding S is not possible.
- $v_1(y_1) = -\infty$: In Stage II-S, concluding I is not possible.

If simpler designs like these give compatible performance characteristics to more complicated designs, the simpler ones are preferred. Therefore, we would like to see how much, if any, simplification is possible. We plan to carry out a computational experiment similar to the ANOVA analysis of Chapter 2 to search for simpler, but equally effective, designs. We intend to study which specification components are influential in determining the characteristics of the resulting designs.

Another topic for future research is the unconditional power function. As we noted in Chapter 4, determination of the power functions is more difficult when there are two sets of hypotheses. Ultimately, we would like to suggest a simpler set of conditions that guarantee monotonicity of the power functions. We anticipate that we can apply the power function analysis for the simpler situation in Chapter 2 to this more complicated problem.

5.4 Estimation From A Two-Stage Adaptive Procedure

As we noted, there have been quite a few papers published in the field of adaptive designs, especially since the mid 1990's. Because our framework encompasses many of the previously proposed procedures, we hope that our methodology gives a new perspective to the field. Our framework gives very powerful tools for designing a two-stage adaptive procedure.

Our framework can control Type I and Type II errors in a statistical hypothesis testing. However, we have not addressed how to draw inferences from the results of the study. Our sense is that the main research in this field would benefit by beginning to shift to the issues concerning estimation when a design permits adaptation. There have been only few papers [e.g., Liu and Chi (2001)] which address the issues in point estimation and confidence intervals for two-stage adaptive procedures. It may be the clinical trial sponsor's interest to obtain an unbiased estimate of Δ, a confidence interval

for Δ using the result in a two-stage adaptive procedure, as well as to provide a "p-value." These issues are not fully investigated yet, and these are not simple problems.

The difficulty of making inference from a two-stage adaptive procedure can be illustrated by the following simple hypothetical example. Suppose that to test $H_0 : \mu_t - \mu_c \leq 0$ against $H_1 : \mu_t - \mu_c > 0$, a sample of size $n_1 = 100$ is taken in Stage I. In the first scenario, suppose that we observe $y_1 = 1.0$, which is greater than $k_2 = .9$. Then the null hypothesis is rejected in Stage I, and the study is terminated. In the second scenario, suppose that $y_1 = .8$. Then we move on to Stage II with sample size of $n_2(.8) = 200$. At the end of Stage II, we observe $y_2 = 2.0$, and suppose that we reject H_0 because $w(.8) = .9$. In both of these two scenarios, H_0 is rejected, but which one gives a stronger evidence against H_0? When only Stage I is considered, scenario 1 gives stronger evidence against H_0. In scenario 2, the sample size of Stage II is twice as large as the one of Stage I, and the observed y_2 is much larger than y_1. How we obtain an unbiased estimate of Δ and a confidence interval for such scenarios is a challenging problem.

It is our intention to continue research on the inference problems mentioned above making use of the perspective provided by our framework. Because our framework covers many previously proposed procedures, if we can establish proper inference procedures in our framework, it would give approaches for many of the already proposed two-stage adaptive procedures.

Appendix

Appendix A

ANOVA Tables For Design Parameter Analyses Of Section 2.5

The following is the detailed outputs of the ANOVA study we conducted for this research. The description of this analysis of the "response surface" is given in Section 2.5. In this appendix, we present an ANOVA table for each of 9 responses we consider. The actual ANOVA table includes 62 different sources (6 main effects, 15 two-way, 20 three-way, 15 four-way and 6 five-way interactions). However all the high-degree interactions are not significant. The ANOVA tables presented here include all the 6 main effects and 2 significant two-way interactions. For each source, DF is the degree of freedom, SS is the sum of squares and Proportion is the SS's proportion to the total sum of squares.

Table A.1: *ANOVA table for Power at* $\lambda = 1/4$.

Source	DF	SS	Proportion
A_0	4	15.1	.006
A_1	4	14.5	.006
f	2	137.5	.057
α_1	2	1364.6	.561
β_1	2	317.2	.130
ξ_1	2	0.0	.000
$f \times \alpha_1$	8	419.2	.172
$f \times \beta_1$	8	90.9	.037
Others	1992	74.5	.031
Total	2024	2433.5	1.000

Table A.2: *ANOVA table for Power at* $\lambda = 1/2$.

Source	DF	SS	Proportion
A_0	4	67.0	.008
A_1	4	61.7	.007
f	2	3.9	.001
α_1	2	3578.4	.425
β_1	2	2554.0	.303
ξ_1	2	0.1	.000
$f \times \alpha_1$	8	1040.8	.124
$f \times \beta_1$	8	819.7	.097
Others	1992	294.3	.035
Total	2024	8419.9	1.000

Table A.3: *ANOVA table for Power at* $\lambda = 3/4$.

Source	DF	SS	Proportion
A_0	4	34.3	.009
A_1	4	26.8	.007
f	2	170.9	.045
α_1	2	814.5	.215
β_1	2	1780.0	.470
ξ_1	2	0.0	.000
$f \times \alpha_1$	8	209.5	.055
$f \times \beta_1$	8	615.4	.162
Others	1992	138.4	.037
Total	2024	3789.8	1.000

Table A.4: *ANOVA table for* R_1 *at* $\lambda = 0$.

Source	DF	SS	Proportion
A_0	4	.368	.026
A_1	4	.306	.022
f	2	3.564	.250
α_1	2	2.694	.189
β_1	2	4.979	.349
ξ_1	2	.006	.000
$f \times \alpha_1$	8	1.340	.094
$f \times \beta_1$	8	.270	.019
Others	1992	.727	.051
Total	2024	14.254	1.000

Table A.5: *ANOVA table for* R_1 *at* $\lambda = 1/4$.

Source	DF	SS	Proportion
A_0	4	.256	.018
A_1	4	.236	.016
f	2	5.404	.371
α_1	2	2.920	.200
β_1	2	3.199	.220
ξ_1	2	.009	.000
$f \times \alpha_1$	8	1.479	.102
$f \times \beta_1$	8	.286	.020
Others	1992	.779	.053
Total	2024	14.568	1.000

Table A.6: *ANOVA table for* R_1 *at* $\lambda = 1/2$.

Source	DF	SS	Proportion
A_0	4	.089	.007
A_1	4	.073	.006
f	2	7.042	.543
α_1	2	1.701	.131
β_1	2	1.102	.085
ξ_1	2	.013	.001
$f \times \alpha_1$	8	1.442	.111
$f \times \beta_1$	8	.867	.067
Others	1992	.641	.049
Total	2024	12.970	1.000

Table A.7: *ANOVA table for* R_1 *at* $\lambda = 3/4$.

Source	DF	SS	Proportion
A_0	4	.244	.019
A_1	4	.039	.003
f	2	7.353	.570
α_1	2	1.050	.081
β_1	2	1.385	.107
ξ_1	2	.011	.001
$f \times \alpha_1$	8	1.003	.078
$f \times \beta_1$	8	1.272	.099
Others	1992	.547	.042
Total	2024	12.904	1.000

Table A.8: *ANOVA table for* R_1 *at* $\lambda = 1$.

Source	DF	SS	Proportion
A_0	4	.539	.042
A_1	4	.116	.009
f	2	5.714	.445
α_1	2	1.849	.144
β_1	2	2.324	.181
ξ_1	2	.000	.000
$f \times \alpha_1$	8	.360	.028
$f \times \beta_1$	8	1.348	.105
Others	1992	.591	.046
Total	2024	12.841	1.000

Table A.9: *ANOVA table for* R_2 *at* $\lambda = 1/4$.

Source	DF	SS	Proportion
A_0	4	.771	.008
A_1	4	.029	.000
f	2	19.703	.200
α_1	2	41.220	.418
β_1	2	12.473	.126
ξ_1	2	.000	.000
$f \times \alpha_1$	8	19.452	.197
$f \times \beta_1$	8	.789	.008
Others	1992	4.28	.043
Total	2024	98.717	1.000

Table A.10: *ANOVA table for* R_2 *at* $\lambda = 1/2$.

Source	DF	SS	Proportion
A_0	4	.316	.009
A_1	4	.001	.000
f	2	8.654	.253
α_1	2	11.956	.349
β_1	2	4.716	.138
ξ_1	2	.000	.000
$f \times \alpha_1$	8	6.727	.196
$f \times \beta_1$	8	.071	.002
Others	1992	1.826	.053
Total	2024	34.267	1.000

Table A.11: *ANOVA table for* R_2 *at* $\lambda = 3/4$.

Source	DF	SS	Proportion
A_0	4	.412	.037
A_1	4	.056	.005
f	2	5.552	.499
α_1	2	1.825	.164
β_1	2	.467	.042
ξ_1	2	.000	.000
$f \times \alpha_1$	8	2.014	.181
$f \times \beta_1$	8	.168	.015
Others	1992	.634	.057
Total	2024	11.128	1.000

Appendix B

Liu's Argument Against "Partially Breaking The Blind"

In the following, we will apply Liu's (2002) argument and show how to calculate the squared difference of the group means in Stage I from the pooled variance and the overall variance from the Stage I. In Appendix B and Appendix C, we use the following same notation. Let x_{igj} be the j^{th} observation from the group g in Stage i, where g takes t if treatment and c if control. The averages are denoted in a usual manner. For example, $\bar{x}_{1t\bullet}$ is the Stage I treatment group mean, and $\bar{x}_{1\bullet\bullet}$ is the mean of all the observations in Stage I. For the variance estimates, we use the following notation. Let s^2_{it}, s^2_{ic} be the Stage i variance estimates from the treatment and control group, respectively. [See (3.3.32), (3.3.33), (3.3.34), (3.3.35).] And let s^2_{ip} be the Stage i pooled variance estimate. [See (3.3.36), (3.3.37).] Finally, let $s^2_{i\bullet}$ be the Stage i overall variance. The overall variance of Stage I is

$$s^2_{1\bullet} = \frac{\sum_{j=1}^{n_{1t}}(x_{1tj} - \bar{x}_{1\bullet\bullet})^2 + \sum_{j=1}^{n_{1c}}(x_{1cj} - \bar{x}_{1\bullet\bullet})^2}{n_{1t} + n_{1c} - 1}. \tag{B.0.1}$$

Suppose that at the end of Stage I, the investigator has an access to the overall sample variance, $s^2_{1\bullet}$. Further suppose that the pooled variance, s^2_{1p}, is also revealed. Then with the above information, the squared difference of the means, $(\bar{x}_{1t\bullet} - \bar{x}_{1c\bullet})^2$, can be calculated.

We have

$$\sum_{j=1}^{n_{1t}}(x_{1tj} - \bar{x}_{1\bullet\bullet})^2 = \sum_{j=1}^{n_{1t}}(x_{1tj} - \bar{x}_{1t\bullet})^2 + n_{1t}(\bar{x}_{1t\bullet} - \bar{x}_{1\bullet\bullet})^2. \tag{B.0.2}$$

Using (B.0.2) and a similar identity for the control group, (B.0.1) can be written as

$$s^2_{1\bullet} = \frac{\sum_{j=1}^{n_{1t}}(x_{1tj} - \bar{x}_{1t\bullet})^2 + n_{1t}(\bar{x}_{1t\bullet} - \bar{x}_{1\bullet\bullet})^2 + \sum_{j=1}^{n_{1c}}(x_{1cj} - \bar{x}_{1c\bullet})^2 + n_{1c}(\bar{x}_{1c\bullet} - \bar{x}_{1\bullet\bullet})^2}{n_{1t} + n_{1c} - 1}.$$

Using (3.3.32) and (3.3.33), $s^2_{1\bullet}$ is

$$s^2_{1\bullet} = \frac{(n_{1t} - 1)s^2_{1t} + (n_{1c} - 1)s^2_{1c} + n_{1t}(\bar{x}_{1t\bullet} - \bar{x}_{1\bullet\bullet})^2 + n_{1c}(\bar{x}_{1c\bullet} - \bar{x}_{1\bullet\bullet})^2}{n_{1t} + n_{1c} - 1}. \tag{B.0.3}$$

Because $\bar{x}_{1\bullet\bullet} = (n_{1t}\bar{x}_{1t\bullet} + n_{1c}\bar{x}_{1c\bullet})/(n_{1t} + n_{1c})$,

$$\bar{x}_{1t\bullet} - \bar{x}_{1\bullet\bullet} = \frac{n_{1c}}{n_{1t} + n_{1c}} (\bar{x}_{1t\bullet} - \bar{x}_{1c\bullet}). \tag{B.0.4}$$

Substituting (*B*.0.4) and a similar result for the control group in (*B*.0.3), we have

$$s^2_{1\bullet} = \frac{1}{n_{1t} + n_{1c} - 1} \left\{ (n_{1t} - 1)s^2_{1t} + (n_{1c} - 1)s^2_{1c} + \frac{n_{1t}n_{1c}}{n_{1t} + n_{1c}} (\bar{x}_{1t\bullet} - \bar{x}_{1c\bullet})^2 \right\}.$$

Moreover, from (3.3.36), we have $(n_{1t} + n_{1c} - 2)s^2_{1p} = (n_{1t} - 1)s^2_{1t} + (n_{1c} - 1)s^2_{1c}$. Therefore,

$$s^2_{1\bullet} = \frac{1}{n_{1t} + n_{1c} - 1} \left\{ (n_{1t} + n_{1c} - 2)s^2_{1p} + \frac{n_{1t}n_{1c}}{n_{1t} + n_{1c}} (\bar{x}_{1t\bullet} - \bar{x}_{1c\bullet})^2 \right\}. \tag{B.0.5}$$

Finally, by solving (*B*.0.5) for $(\bar{x}_{1t\bullet} - \bar{x}_{1c\bullet})^2$, we have

$$(\bar{x}_{1t\bullet} - \bar{x}_{1c\bullet})^2 = \frac{n_{1t} + n_{1c}}{n_{1t}n_{1c}} \left\{ (n_{1t} + n_{1c} - 1)s^2_{1\bullet} - (n_{1t} + n_{1c} - 2)s^2_{1p} \right\}. \tag{B.0.6}$$

Therefore, with the knowledge of $s^2_{1\bullet}$ and s^2_{1p}, we can calculate $(\bar{x}_{1t\bullet} - \bar{x}_{1c\bullet})^2$ When $n_{1t} = n_{1c} \equiv n_1$, (*B*.0.6) simplifies to

$$(\bar{x}_{1t\bullet} - \bar{x}_{1c\bullet})^2 = \frac{4n_1 - 2}{n_1} s^2_{1\bullet} - \frac{4n_1 - 4}{n_1} s^2_{1p}.$$

When we do not assume that the population variances are equal, the individual group's sample variances, s^2_{1t} and s^2_{1c}, need to be revealed instead of the pooled estimate, s^2_{1p}. Nonetheless, we can calculate s^2_{1p} using (3.3.36) to get the same result.

Appendix C

The Ratio Of Variances

We consider the ratio of variances (standard deviations) as introduced in Section 3.3.2. The goal is to show that when the sample sizes for one stage are at least moderate, the pooled variance from that stage and the pooled variance utilizing entire data set from both stages are similar. In order to avoid algebraic complexity, we assume that the sample sizes from the treatment and control groups are equal. We denote the Stage I sample size by n_1 and the Stage II sample size by n_2.

The variance estimate from the treatment group utilizing both stages can be written as

$$s^2_{\bullet t} = \frac{1}{n^*} \sum_{i=1}^{2} \sum_{j=1}^{n_i} (x_{itj} - \bar{x}_{\bullet t\bullet})^2,$$

where $n^* = n_1 + n_2 - 1$. And we can write the variance estimate from the treatment group for each Stage as follows:

$$s^2_{it} = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (x_{itj} - \bar{x}_{it\bullet})^2,$$

for $i = 1, 2$. Because $\bar{x}_{\bullet t\bullet} = (n_1 \bar{x}_{1t\bullet} + n_2 \bar{x}_{2t\bullet})/(n_1 + n_2)$, we can write

$$\bar{x}_{1t\bullet} - \bar{x}_{\bullet t\bullet} = \frac{n_2}{n_1 + n_2} (\bar{x}_{1t\bullet} - \bar{x}_{2t\bullet}), \qquad \bar{x}_{2t\bullet} - \bar{x}_{\bullet t\bullet} = \frac{n_1}{n_1 + n_2} (\bar{x}_{2t\bullet} - \bar{x}_{1t\bullet}).$$

Using these identities, $n^*\, s^2_{\bullet t}$ is expressed as

$$\begin{aligned} n^*\, s^2_{\bullet t} &= \sum_{j=1}^{n_1} (x_{1ti} - \bar{x}_{1t\bullet})^2 + n_1 (\bar{x}_{1t\bullet} - \bar{x}_{\bullet t\bullet})^2 + \sum_{j=1}^{n_2} (x_{2ti} - \bar{x}_{2t\bullet})^2 + n_2 (\bar{x}_{2t\bullet} - \bar{x}_{\bullet t\bullet})^2 \\ &= (n_1 - 1) s^2_{1t} + (n_2 - 1) s^2_{2t} + \frac{n_1 n_2}{n_1 + n_2} \left(\bar{x}_{1t\bullet} - \bar{x}_{2t\bullet}\right)^2. \end{aligned} \tag{C.0.1}$$

Similarly, for the control group, we have

$$n^*\, s^2_{\bullet c} = (n_1 - 1) s^2_{1c} + (n_2 - 1) s^2_{2c} + \frac{n_1 n_2}{n_1 + n_2} \left(\bar{x}_{1c\bullet} - \bar{x}_{2c\bullet}\right)^2.$$

Because $s_{1p}^2 = (s_{1t}^2 + s_{1c}^2)/2$, $s_{2p}^2 = (s_{2t}^2 + s_{2c}^2)/2$ and $s_{\bullet p}^2 = (s_{\bullet t}^2 + s_{\bullet c}^2)/2$, we have

$$n^* \, s_{\bullet p}^2 = (n_1 - 1)s_{1p}^2 + (n_2 - 1)s_{2p}^2 + \frac{n_1 n_2}{n_1 + n_2} \frac{(\bar{x}_{1t\bullet} - \bar{x}_{2t\bullet})^2 + (\bar{x}_{1c\bullet} - \bar{x}_{2c\bullet})^2}{2}.$$

Therefore,

$$\begin{aligned} n^* \frac{s_{\bullet p}^2}{s_{1p}^2} &= (n_1 - 1) + (n_2 - 1)\frac{s_{2p}^2}{s_{1p}^2} + \frac{n_1 n_2}{n_1 + n_2} \frac{(\bar{x}_{1t\bullet} - \bar{x}_{2t\bullet})^2 + (\bar{x}_{1c\bullet} - \bar{x}_{2c\bullet})^2}{2s_{1p}^2} \\ &= (n_1 - 1) + \frac{(n_2 - 1)\dfrac{s_{2p}^2}{\sigma^2} + \dfrac{n_1 n_2}{2(n_1 + n_2)} \left\{ \dfrac{(\bar{x}_{1t\bullet} - \bar{x}_{2t\bullet})^2 + (\bar{x}_{1c\bullet} - \bar{x}_{2c\bullet})^2}{\sigma^2} \right\}}{s_{1p}^2/\sigma^2} \\ &= (n_1 - 1) + \frac{U_2/2 + U_3/2}{U_1/2(n_1 - 1)}, \end{aligned}$$

where

$$\begin{aligned} U_1 &= 2(n_1 - 1)s_{1p}^2/\sigma^2 \\ U_2 &= 2(n_2 - 1)s_{2p}^2/\sigma^2 \\ U_3 &= \frac{n_1 n_2}{n_1 + n_2} \left\{ \frac{(\bar{x}_{1t\bullet} - \bar{x}_{2t\bullet})^2 + (\bar{x}_{1c\bullet} - \bar{x}_{2c\bullet})^2}{\sigma^2} \right\}. \end{aligned}$$

We have that $U_1 \sim \chi^2_{2(n_1-1)}$ and $U_2 \sim \chi^2_{2(n_2-1)}$. For U_3, we note that both $\bar{x}_{1t\bullet} - \bar{x}_{2t\bullet}$ and $\bar{x}_{1c\bullet} - \bar{x}_{2c\bullet}$ are distributed as Normal $\big(0, \, \sigma^2(n_1 + n_2)/n_1 n_2\big)$, and they are independent of each other. Thus $U_3 \sim \chi^2_2$. Moreover, U_1, U_2 and U_3 are mutually independent.

Let $U_4 = U_2 + U_3$. Then we can write

$$n^* \frac{s_{\bullet p}^2}{s_{1p}^2} = (n_1 - 1) + \frac{U_4/2n_2}{U_1/2(n_1 - 1)} \times n_2$$

Because U_2 and U_3 are independent χ^2 random variables with respective degrees of freedom, $2(n_2-1)$ and 2, $U_4 \sim \chi^2_{2n_2}$, which is independent of U_1. Let $F = \dfrac{U_4/2n_2}{U_1/2(n_1 - 1)}$. Then $F \sim F_{2n_2,\, 2(n_1-1)}$. Finally,

$$\frac{s_{\bullet p}^2}{s_{1p}^2} = \frac{n_1 - 1}{n_1 + n_2 - 1} + \frac{F}{n_1 + n_2 - 1} \times n_2. \tag{C.0.2}$$

Therefore, given n_1 and n_2 we can find a 95% prediction interval for $s_{\bullet p}/s_{1p}$ using (C.0.2). The following Table C.1 summarizes the prediction intervals for various n_1 and n_2. As the table shows, when Stage I is fairly large, the estimate of standard error from Stage I is very similar to the estimate utilizing both stages.

When we do not assume that the underlying population variances for the treatment and control groups are equal, we need to estimate each population variance using the data from just one group,

Table C.1: *Ratio of pooled standard errors, $s_{\bullet p}/s_{1p}$ for various n_1 and n_2.*

	n_2				
n_1	50	100	150	200	400
50	(.91, 1.12)	(.90, 1.13)	(.89, 1.14)	(.89, 1.15)	(.88, 1.15)
100	(.95, 1.06)	(.94, 1.08)	(.93, 1.08)	(.93, 1.09)	(.92, 1.10)
150	(.96, 1.04)	(.95, 1.06)	(.95, 1.06)	(.94, 1.07)	(.94, 1.07)
200	(.97, 1.03)	(.96, 1.04)	(.96, 1.05)	(.95, 1.05)	(.95, 1.06)
400	(.98, 1.02)	(.98, 1.02)	(.98, 1.03)	(.97, 1.03)	(.97, 1.04)

treatment or control. In other words, we cannot use the pooled estimates. Suppose that in Stage I, a sample of size n_{1t} is taken from the treatment group, and in Stage II, a sample of size n_{2t} is taken from the treatment group. Then similarly to (C.0.1), we have

$$(n_{1t}-1)s^2_{\bullet t} = (n_{1t}-1)s^2_{1t} + (n_{2t}-1)s^2_{2t} + \frac{n_{1t}n_{2t}}{n_{1t}+n_{2t}}(\bar{x}_{1t\bullet} - \bar{x}_{2t\bullet})^2.$$

And a similar argument as before leads to

$$\frac{s^2_{\bullet t}}{s^2_{1t}} = \frac{n_{1t}-1}{n_{1t}+n_{2t}-1} + \frac{F}{n_{1t}+n_{2t}-1} \times n_{2t},$$

where $F \sim F_{n_{2t}, n_{1t}-1}$.

Tabel C.2 summarizes the ratio of standard errors for this case.

Table C.2: *Ratio of nonpooled standard errors, $s_{\bullet t}/s_{1t}$ for various n_{1t} and n_{2t}.*

	n_{2t}				
n_{1t}	50	100	150	200	400
50	(.88, 1.18)	(.87, 1.20)	(.86, 1.21)	(.85, 1.22)	(.84, 1.23)
100	(.93, 1.10)	(.91, 1.12)	(.91, 1.13)	(.90, 1.13)	(.89, 1.14)
150	(.95, 1.07)	(.94, 1.08)	(.93, 1.09)	(.92, 1.10)	(.91, 1.11)
200	(.96, 1.05)	(.95, 1.06)	(.94, 1.07)	(.94, 1.08)	(.93, 1.09)
400	(.98, 1.03)	(.97, 1.03)	(.97, 1.04)	(.96, 1.04)	(.95, 1.05)

Appendix D

Splus Functions

In this research, we use Splus to perform computations. We use multiple short functions, rather than a few lengthy ones. In the following, the main functions are listed with short descriptions of what they are used for. The inputs of a function are usually named intuitively, i.e., `sigma` is σ, `alpha` is α, and so on.

D.1 Preliminaries And Stage I

The following two functions convert $a = 1 - \Phi(z_a)$ to z_a and vice versa.

```
atoz_function(a){qnorm(1-a)}   ztoa_function(z){1-pnorm(z)}
```

The following simply redefine Splus functions to save spaces.

```
an_function(x){as.numeric(x)} ; av_function(x){as.vector(x)}
rr_function(x, r){round(x, r)} ; r0_function(x){rr(x, 0)}
r1_function(x){rr(x, 1)} ; r2_function(x){rr(x, 2)}
r3_function(x){rr(x, 3)} ; r4_function(x){rr(x, 4)}
```

The function "`calcn`" calculates the conventional one-stage sample size given σ, μ_0, μ_1, α and β.

```
calcn_function(sigma, mu0, mu1, alpha, beta){
   (sqrt(2)*sigma)^2*(atoz(alpha)-atoz(1-beta))^2/(mu1-mu0)^2}
```

The function "`findk`" calculates the stage I critical values, k_1 or k_2. The inputs are σ, α or $1 - \beta$, μ (μ_0 for α and μ_1 for β) and n_1.

```
findk_function(sigma, ab, mu, n1){
   r4(atoz(ab)/sqrt(n1)+mu/(sqrt(2)*sigma))}
```

The next function, `step1`, calculates the specification components of Stage I. The sample size n_1 must be given, either indirectly through N and f, or directly. If n_1 is specified elsewhere, it should

be given as an input, otherwise, let `n1p=0`. The critical value, k_1 or k_2, are calculated using n_1, α_1 and/or β_1. However, they can be specified elsewhere, too. Then α_1 and/or β_1 are computed in this function using n_1 and k_1 and/or k_2. When k_1 or k_2 are not specified, these input should be `k1p=0 k2p=0`.

```
step1_function(sigma, mu0, mu1, f, alpha, beta, alpha1, beta1,
               n1p=0, k1p=0, k2p=0){
   N_ceiling(calcn(sigma, mu0, mu1, alpha, beta)) ; n1x_round(N*f)
   n1x_c(n1p, n1x)[an(n1p==0)+1]
   k1x_findk(sigma, 1-beta1, mu1, n1x) ; k2x_findk(sigma, alpha1, mu0, n1x)
   k1x_c(k1p, k1x)[an(k1p==0)+1] ; k2x_c(k2p, k2x)[an(k2p==0)+1]
   c(N, n1x, k1x, k2x)}
```

The following `acr` (which stands for accept, continue, reject) calculates the Stage I probabilities given σ, μ, n_1, k_1 and k_2. If `mu` is a vector, this function calculates the Stage I probabilities for each value of `mu`.

```
acr_function(sigma, mu, n1, k1, k2){
   xx_mu/(sqrt(2)*sigma)
   ss_1/sqrt(n1)
   Rej_r3(1-pnorm(k2, xx, ss)) ; Acc_r3(pnorm(k1, xx, ss)) ; Con_1-Rej-Acc
   cbind(Acc, Con, Rej)}
```

The following two functions are used in the necessary integrations in obtaining A-functions.

```
ins1_function(yt, sigmat, mut, k1t, pt, n1t){
   xxt_mut/(sqrt(2)*sigmat) ; sst_1/sqrt(n1t)
   (yt-k1t)^pt*dnorm(yt, xxt, sst)}


out1_function(sigma, mu, p, n1, k1, k2){#Integration for calculating a1
   integrate(ins1, lower=k1, upper=k2, sigmat=sigma, mut=mu,
             k1t=k1, pt=p, n1t=n1)$integral}
```

The next function, "`step2`," calculates the value of the flat conditional power function. The output of this function is then used to decide the starting value $A(k_1, \xi)$ for the functions with linear, quadratic and square-root shapes.

```
step2_function(sigma, mu, pow2, n1, k1, k2){
```

```
   stg1_acr(sigma, mu, n1, k1, k2)
   pcon_an(stg1[1,2])
   an(r4(pow2/pcon))}
```

Finally, "step3" solves a_1 in $A(y_1, \xi) = a_0 + a_1(y_1 - k_1)^r$. The input a0 is a_0 and pow2 is the unconditional power for Stage II at $\mu =$ mu. The output are a_0, a_1 for $r = 1, 2, .5$, and the ending point of the conditional power function, i.e., $A(k_2, \xi)$, which needs to be below 1.

```
step3_function(sigma, mu, a0, pow2, n1, k1, k2){

   xi_mu/(sqrt(2)*sigma) ; sss_1/sqrt(n1)
   stg1_acr(sigma, mu, n1, k1, k2)
   pcon_an(stg1[1,2])
   ome_out1(sigma, mu, 1,  n1, k1, k2)
   sqr_out1(sigma, mu, 2,  n1, k1, k2)
   roo_out1(sigma, mu, .5, n1, k1, k2)

   a1L_r4((pow2-a0*pcon)/ome) ; a1Le_r4(a0+a1L*(k2-k1))
   a1S_r4((pow2-a0*pcon)/sqr) ; a1Se_r4(a0+a1S*(k2-k1)^2)
   a1R_r4((pow2-a0*pcon)/roo) ; a1Re_r4(a0+a1R*sqrt(k2-k1))

   FL_an(r4(pow2/pcon))

   amat_cbind(a0, a1L, a1Le, a0, a1S, a1Se, a0, a1R, a1Re, FL)
   amat}
```

D.2 Stage II Conditional Powers And Sample Sizes

In this section, we present functions that are used to design Stage II. The first of these is Afs, which calculates the actual value of $A(y_1, \xi)$. Note that there are 2 k_1's and 2 k_2's. For "missing observations" and "unknown variance," Stage I critical values k_1 and k_2 are updated. For this function, we can specify the original k values and the updated k values in ascending order, k1x $<$ k1p $<$ k2p $<$ k2x. The values of $A(y_1, \xi)$ in between k1x and k1p, and in between k2p and k2x are constant. The other inputs are a_0, a_1 and r which specify $A(y_1, \xi) = a_0 + a_1(y_1 - k_1)^r$. The last input const is the constant by which $A(y_1, \xi)$ is multiplied. If the Stage I critical values are not modified, const should be 1.

```
Afs_function(y1, k1x, k1p, k2p, k2x, ab0, ab1, pab, const){
   sm_an(y1<k1p) ; bg_an(y1>k2p)
   pos_sm-bg # 1 for small, -1 for big, 0 otherwise
   y1_an(pos == 1) * k1p + an(pos==0)*y1 + an(pos==-1) * k2p
   #For extension regions, y1 is equal to k1p or k2p.
   temp_const * (ab0 + ab1 * (y1 - k1p)^pab)}
```

The following series of functions whose name starts with `n2` are used to specify a shape of $n_2(y_1)$ function. First, `n2original` simply calculates the Stage II sample size function from $A(y_1, \xi_0)$ and $A(y_1, \xi_1)$. The new input, `a0cA` is the η that adjusts $A(y_1, \xi_0)$ when the Stage I critical values are modified. The next function, `n2check` checks where the maximum and minimum of the sample size function occur. If the maximum occur somewhere other than at k_1, or if the minimum occur somewhere other than at k_2, it indicates that the Stage II sample size function is not nonincreasing. The output of this function is the value of y_1 where the maximum and minimum Stage II sample size occur. The third function, `n2dec` gives a monotone Stage II sample size function. If, for example, the maximum occurs at $y_1 = t$, $n_2(y_1)$ for all y_1 in (k_1, t) will be equal to $n_2(t)$. Finally, `n2func` gives the final $n_2(y_1)$. It adds $\eta =$ `addton2` to the basic shape, which is given by `n2dec`. It also checks if the resulting sample size is within the preset maximum (`n2max`) and minimum (`n2min`). If the sample size function is outside of this range, the sample size is truncated to `n2max` or increased to `n2min`.

```
n2original_function(y1, sigma, mu0, mu1, k1x, k1p, k2p, k2x,
                    a0, a1, pa, a0cA, b0, b1, pb){
   za0_atoz(Afs(y1, k1x, k1p, k2p, k2x, a0, a1, pa, a0cA))
   za1_atoz(Afs(y1, k1x, k1p, k2p, k2x, b0, b1, pb, 1))
   2*(sigma*(za0-za1)/(mu1-mu0))^2}
```

```
n2check_function(sigma, mu0, mu1, k1x, k1p, k2p, k2x,
                 a0, a1, pa, a0cA, b0, b1, pb){
   y1_seq(k1x, k2x, length=200)
   n2_n2original(y1, sigma, mu0, mu1, k1x, k1p, k2p, k2x,
                 a0, a1, pa, a0cA, b0, b1, pb, a0cC)
   maxpos_min(seq(1, length(y1))[n2==max(n2)])
```

```
   minpos_max(seq(1, length(y1))[n2==min(n2)])
   y1top_y1[maxpos] ; y1btm_y1[minpos]
   c(y1top, y1btm)}


n2dec_function(y1, sigma, mu0, mu1, k1x, k1p, k2p, k2x,
               a0, a1, pa, a0cA, b0, b1, pb, a0cC, y1top, y1btm){
   sm_an(y1<y1top) ; bg_an(y1>y1btm) ;
   pos_sm-bg # 1 for small, -1 for big, 0 otherwise
   y1_an(pos==1)*y1top + an(pos==0)*y1 + an(pos==-1)*y1btm
   n2original(y1, sigma, mu0, mu1, k1x, k1p, k2p, k2x,
   a0, a1, pa, a0cA, b0, b1, pb)}


n2func_function(y1, sigma, mu0, mu1, k1x, k1p, k2p, k2x,
                a0, a1, pa, a0cA, b0, b1, pb, a0cC, n2max, n2min){
   check_n2check(sigma, mu0, mu1, k1x, k1p, k2p, k2x,
                 a0, a1, pa, a0cA, b0, b1, pb)
   y1top_check[1] ; y1btm_check[2]
   n2_n2dec(y1, sigma, mu0, mu1, k1x, k1p, k2p, k2x,
            a0, a1, pa, a0cA, b0, b1, pb, a0cC, y1top, y1btm)
   n2_n2+addton2
   toosmall_an(n2<n2min) ; toolarge_an(n2>n2max)
   code_toolarge-toosmall #This is 1, 0 or -1
   n2t_n2max*an(code==1)+n2*an(code==0)+n2min*an(code==-1)
   n2t}
```

The following function, `A0n2wA1` finds $A(y_1, \xi^*)$ and $w(y_1)$ given the inputs $A(y_1, \xi_0)$ and $n_2(y_1)$. Because `n2` is obtained through the function, `n2func`, in which $A(y_1, \xi_0)$ and $A(y_1, \xi^*)$ is used to find the original $n_2(y_1)$, the initial $A(y_1, \xi^*)$ must be given. The new $n_2(y_1)$ accommodates restrictions on the maximum and minimum and the restrictions on the shape. Then using this new $n_2(y_1)$ and $A(y_1, \xi_0)$, this `A0n2wA1` gives the resulting $A(y_1, \xi^*)$ and $w(y_1)$.

```
A0n2wA1_function(y1, sigma, mu0, mu1, k1x, k1p, k2p, k2x,
                 a0, a1, pa, a0cA, b0, b1, pb, a0cC,
```

```
                n2max, n2min, n1){
   xi0_mu0/(sqrt(2)*sigma) ; xi1_mu1/(sqrt(2)*sigma) ; ss_1/sqrt(n1)
   const_c(a0cC, a0cA)[an(a0cA==a0cC)+1]
   A0_Afs(y1, k1x, k1p, k2p, k2x, a0, a1, pa, const) ; za0_atoz(A0)
   n2_n2func(y1, sigma, mu0, mu1, k1x, k1p, k2p, k2x,
             a0, a1, pa, a0cA, b0, b1, pb, a0cC, n2max, n2min)
   A1_ztoa(za0-sqrt(n2)*(xi1-xi0))
   w_za0/sqrt(n2)+xi0
   list(A0, A1, n2, w)}
```

The following `PLOTall` simply plots $A(y_1, \xi_0)$, $A(y_1, \xi^*)$, $n_2(y_1)$ and $w(y_1)$.

```
PLOTall_function(sigma, mu0, mu1, k1x, k1p, k2p, k2x,
                 a0, a1, pa, a0cA, b0, b1, pb, a0cC,
                 n2max, n2min, n1){
   y1_seq(k1x, k2x, length=400)
   temp_A0n2wA1(y1, sigma, mu0, mu1, k1x, k1p, k2p, k2x,
                a0, a1, pa, a0cA, b0, b1, pb, a0cC, n2max, n2min, n1)
   A0_temp[[1]] ; A1_temp[[2]] ; n2_temp[[3]] ; w_temp[[4]]
   plot(c(k1x, k2x), c(0,1), xlab="", ylab="", type="n")
   lines(y1, A0, lty=1, lwd=2)
   lines(y1, A1, lty=1, lwd=2)
   plot(c(k1x, k2x), c(0, max(n2)), xlab="", ylab="", type="n")
   lines(y1, n2, lty=1, lwd=2)
   plot(c(k1x, k2x), c(min(w), max(w)), xlab="", ylab="", type="n")
   lines(y1, w, lty=1, lwd=2)
   list(y1, A0, A1, n2, w)}
```

D.3 Calculating The Power And Expected Sample Size

The functions in this subsection are used to calculate the power and expected sample sizes of a procedure. The funciton `condpow` calculates the conditional power at given $\mu =$ `mu`. And `integpow` and `pow` perform integration to calculate the unconditional power. And `integn2` computes the sample size for Stage II for a given y_1, and `en2n` performs integration to calculate the expected sample size for Stage II for a given $\mu =$ `mu`. Finally, `powV2` calculates the unconditional power and expected

sample sizes for various values of $\mu =$ `mu`. The output of `powV2` is a table that summarizes the Stage I probabilities, Stage II probabilities and expected sample sizes.

```
condpow_function(y1, sigma, mu0, mu1, k1x, k1p, k2p, k2x,
                 a0, a1, pa, a0cA, b0, b1, pb, a0cC,
                 n2max, n2min, n1, mu){
   temp_A0n2wA1(y1, sigma, mu0, mu1, k1x, k1p, k2p, k2x,
                a0, a1, pa, a0cA, b0, b1, pb, a0cC,
                n2max, n2min, n1)
   n2_temp[[3]] ; w_temp[[4]]
   xi_mu/(sqrt(2)*sigma) ; ss_1/sqrt(n2)
   1-pnorm(w, xi, ss)}

integpow_function(y1t, sigmat, mu0t, mu1t, k1xt, k1pt, k2pt, k2xt,
                  a0t, a1t, pat, a0cAt, b0t, b1t, pbt, a0cCt,
                  n2maxt, n2mint, n1t, mut){
   cond_condpow(y1t, sigmat, mu0t, mu1t, k1xt, k1pt, k2pt, k2xt,
             a0t, a1t, pat, a0cAt, b0t, b1t, pbt, a0cCt,
             n2maxt, n2mint, n1t, mut)
   dnor_dnorm(y1t, mut/(sqrt(2)*sigmat), 1/sqrt(n1t))
   cond*dnor}

pow_function(sigma, mu0, mu1, k1x, k1p, k2p, k2x,
             a0, a1, pa, a0cA, b0, b1, pb, a0cC, n2max, n2min, n1, mu){
   st2_integrate(integpow, lower=k1x, upper=k2x, sigmat=sigma, mu0t=mu0,
                 mu1t=mu1, k1xt=k1x, k1pt=k1p, k2pt=k2p, k2xt=k2x,
                 a0t=a0, a1t=a1, pat=pa, a0cAt=a0cA,
                 b0t=b0, b1t=b1, pbt=pb, a0cCt=a0cC,
                 n2maxt=n2max, n2mint=n2min, n1t=n1, mut=mu)$integral
   st1_1-pnorm(k2x, mu/(sqrt(2)*sigma), 1/sqrt(n1))
   st1_r3(st1) ; st2_r3(st2)
   c(st1, st2, st1+st2)}

integn2_function(y1t, sigmat, mu0t, mu1t, k1xt, k1pt, k2pt, k2xt,
                 a0t, a1t, pat, a0cAt, b0t, b1t, pbt, a0cCt,
                 n2maxt, n2mint, n1t, mut){
```

```
   n2_n2func(y1t, sigmat, mu0t, mu1t, k1xt, k1pt, k2pt, k2xt,
             a0t, a1t, pat, a0cAt, b0t, b1t, pbt, a0cCt,
             n2maxt, n2mint)
   dnor_dnorm(y1t, mut/(sqrt(2)*sigmat), 1/sqrt(n1t))
   n2*dnor}

en2n_function(sigma, mu0, mu1, k1x, k1p, k2p, k2x,
              a0, a1, pa, a0cA, b0, b1, pb, a0cC, n2max, n2min, n1, mu){
   en2_integrate(integn2, lower=k1x, upper=k2x,
                 sigmat=sigma, mu0t=mu0, mu1t=mu1,
                 k1xt=k1x, k1pt=k1p, k2pt=k2p, k2xt=k2x,
                 a0t=a0, a1t=a1, pat=pa, a0cAt=a0cA,
                 b0t=b0, b1t=b1, pbt=pb, a0cCt=a0cC,
                 n2maxt=n2max, n2mint=n2min, n1t=n1, mut=mu)$integral
   r1(en2)}

pow2V_function(sigma, mu0, mu1, k1x, k1p, k2p, k2x,
               a0, a1, pa, a0cA, b0, b1, pb, a0cC,
               n2max, n2min, n1, stm, enm, byby){
   Mu_seq(stm, enm, byby) ; len_length(Mu)
   xiv_Mu/(sqrt(2)*sigma) ; ss_1/sqrt(n1)
   Acc1_r3(pnorm(k1x, xiv, ss))
   Con1_r3(pnorm(k2x, xiv, ss)-pnorm(k1x, xiv, ss))
   st2_matrix(0, nrow=len, ncol=3)
   for(i in 1:len){st2[i,]_pow(sigma, mu0, mu1, k1x, k1p, k2p, k2x,
                               a0, a1, pa, a0cA, b0, b1, pb, a0cC,
                               n2max, n2min, n1, Mu[i])}
   en2_rep(0, len)
   for(i in 1:len){en2[i]_en2n(sigma, mu0, mu1, k1x, k1p, k2p, k2x,
                               a0, a1, pa, a0cA, b0, b1, pb, a0cC,
                               n2max, n2min, n1, Mu[i])}
   Rej1_st2[,1] ; Rej2_st2[,2] ; Rej_st2[,3]
   EN2_en2 ; ENT_EN2+n1
   cbind(Mu, Acc1, Con1, Rej1, Rej2, Rej, EN2, ENT)}
```

D.4 The Example Of Section 3.2.5

In this section, we show how the functions introduced above are actually used to construct the example of Section 3.2.5. First, the following constants and parameters are defined.

```
mu0_0 ; mu1_1 ; sigmat_5 ; sigmac_3
alpha_.025 ; beta_.1 ; alpha1_.005 ; beta1_.05
lamb0_.5
rho_1-beta
```

The Stage I sample size can be calculated using `calcn` with `sigma` $= a$. In place of `sigma`, we use the following `a`.

```
a_r4(sqrt((lamb0*sigmat^2+sigmac^2)/2))
```

So N_c is

```
Nc_r0(calcn(a, mu0, mu1, alpha, beta))
```

Because we take 75% of N_c in Stage I, we have

```
n1c_r0(Nc*.75) ; m1_n1c
```

As discussed in Section 3.2.2, we can use this `m1` as the Stage I "sample size." The Stage I critical values are calculated as follows:

```
k1_r4(findk(a, 1-beta1, mu1, m1))
k2_r4(findk(a, alpha1, mu0, m1))
```

We will need α_2, ρ_1 and ρ_2, and they are given by

```
alpha2_alpha-alpha1
rho1_an(acr(a, mu1, n1, k1, k2)[,3])
rho2_rho-rho1
```

To find $A(y_1, \xi_0)$ and $A(y_1, \xi_1)$, we first decide the starting values, i.e., `a0` $= A(k_1, \xi_0)$ and `b0` $= A(k_1, \xi_1)$. The function `step2` is used to calculate the values of the flat conditional power function. We have

```
> step2(a, c(mu0, mu1), c(alpha2, rho2), m1, k1, k2)
       [,1]
[1,] 0.1709
[2,] 0.8599
```

Therefore, a0 and b0 must be smaller than .1709 and .8599, respectively. We choose to use a0=.08 and b0=.80. Then step3 is used to find appropriate a_1 and b_1 for the conditional power functions at $\xi = \xi_0$ and $\xi = \xi_1$.

```
> step3(sigma, mu0, .08, alpha2, m1, k1, k2)
       a0   a1L   a1Le   a0     a1S   a1Se   a0    a1R   a1Re     FL
[1,] 0.08 2.699 0.3718 0.08 49.8382 0.6624 0.08 0.5433 0.2586 0.1709
> step3(sigma, mu1, .8, rho2, m1, k1, k2)
      a0    a1L   a1Le  a0     a1S   a1Se  a0   a1R   a1Re     FL
[1,] 0.8 0.9233 0.8998 0.8 11.8459 0.9384 0.8 0.244 0.8802 0.8599
```

Because the quadratic form is used both for $A(y_1, \xi_0)$ and for $A(y_1, \xi_1)$ in this example, we have

```
pa_2 ; pb_2
a1_49.8382 ; b1_11.8459
```

Note that pa and pb are r in $A(y_1, \xi) = a_0 + a_1(y_1 - k_1)^r$. To graph $A(y_1, \xi_0)$ and $A(y_1, \xi_1)$, $m_2(y_1)$ and $w(y_1)$, we use PLOTall. First, we set the maximum and minimum of $m_2(y_1)$.

```
n2max_124 ; n2min_0
pow2PLOT(a, mu0, mu1, k1, k1, k2, k2,
         a0, a1, pa, 1, b0, b1, pb, 1,
         n2max, n2min, m1)
```

Because at this planning stage, there is no missing observations, we use k1 for the inputs k1p and k1x, k2 for k2p and k2x, and also we use 1 for both a0cA and a0cC. Finally,

```
pow2V(a, mu0, mu1, k1, k1, k2, k2,
      a0, a1, pa, 1, b0, b1, pb, 1,
      n2max, n2min, m1, stm=0, enm=1.25, byby=.25)
```

gives the following table.

```
       Mu  Acc1  Con1  Rej1  Rej2   Rej  EN2   ENT
[1,] 0.00 0.878 0.117 0.005 0.020 0.025  9.8 179.8
[2,] 0.25 0.678 0.291 0.031 0.098 0.129 22.7 192.7
[3,] 0.50 0.405 0.474 0.121 0.255 0.376 33.7 203.7
[4,] 0.75 0.173 0.507 0.320 0.369 0.689 32.5 202.5
[5,] 1.00 0.050 0.357 0.593 0.307 0.900 20.5 190.5
[6,] 1.25 0.009 0.165 0.826 0.154 0.980  8.5 178.5
```

which is presented in Table 3.1.

Then because of missing observations, we have $n'_{1t} = 296$ and $n'_{1c} = 141$. We let `n1tA=296` and `n1cA=141` and calculate the λ'_1 as `lamb1=n1cA/n1tA`, which yields `lamb1=.4764`. Then m'_1 is 145.0 by

```
m1A_r1(n1cA*(lamb0*sigmat^2+sigmac^2)/(lamb1*sigmat^2+sigmac^2)).
```

The updated Stage I critical values, k'_1 and k'_2 are

```
k1A_r4(sqrt(m1)/sqrt(m1A) * (k1-xi1) + xi1)
k2A_r4(sqrt(m1)/sqrt(m1A) * (k2-xi0) + xi0)
```

To find the correct θ to use with $\theta A(y_1, \xi_0)$, we first let $\theta = 1$ and calculate the resulting Type I error rate.

```
pow2V(a, mu0, mu1, k1A, k1, k2, k2A,
      a0, a1, pa, 1, b0, b1, pb, 1,
      n2max, n2min, m1A, stm=0, enm=1.25, byby=.25)
```

yields

```
       Mu  Acc1  Con1  Rej1  Rej2   Rej  EN2   ENT
[1,] 0.00 0.830 0.165 0.005 0.029 0.034 14.1 159.1
[2,] 0.25 0.619 0.354 0.027 0.121 0.148 27.8 172.8
[3,] 0.50 0.365 0.535 0.101 0.294 0.395 37.9 182.9
[4,] 0.75 0.160 0.576 0.265 0.426 0.691 36.0 181.0
[5,] 1.00 0.050 0.442 0.508 0.385 0.893 24.1 169.1
[6,] 1.25 0.011 0.241 0.748 0.227 0.975 11.5 156.5
```

Then we can calculate the correct multiplier, θ to be .020/.029 = .6897. Finally, by trial and error approach, we find that $\eta = 33$ gives the desired power at the original alternative. We need to define this $\eta =$ `addton2` outside of the main functions as:

```
addton2_33
```

Then we reconstruct the above table using

```
pow2V(a, mu0, mu1, k1A, k1, k2, k2A,
      a0, a1, pa, 1, b0, b1, pb, .6897,
      n2max, n2min, m1A, stm=0, enm=1.25, byby=.25)
```

which produces Table 3.2.

Bibliography

Bibliography

Aras, G. (2001), Superiority, noninferiority, equivalence, and bioequivalence – revisited, *Drug Information Journal*, **35**, 1157-1164.

Bauer, P. and Köhne, L. K. (1994), Evaluation of experiments with adaptive interim analyses, *Biometrics*, **50**, 1029-1041. Correction in *Biometrics*, **52**, 380.

Brannath, W., Bauer, P., Maurer, W. and Posch, M. (2003), Sequential tests for noninferiority and superiority, *Biometrics*, **59**, 106-114.

Cui, L., Hung, J. and Wang, S. J. (1999), Modification of sample size in group sequential clinical trials, *Biometrics*, **55**, 853-857.

CPMP (2000), Points to consider on switching between superiority and non-inferiority, (http://www.emea.eu.int/pdfs/human/ewp/048299en.pdf).

Denne, J. S. (2001), Sample size recalculation using conditional power, *Statistics in Medicine*, **20**, 2645-2660.

Gould, A. L. and Shih, W. J. (1998), Modifying the design of ongoing trials without unblinding, *Statistics in Medicine*, **17**, 89-100.

Gould, A. L. (2001), Sample size re-estimation: recent developments and practical considerations, *Statistics in Medicine*, **20**, 2625-2643.

Hwang, I. K. and Morikawa, T. (1999), Design issues in noninferiority/equivalence trials, *Drug Information Journal*, **33**, 1205-1218.

Hommel, G. and Kropf, S. (2001), Clinical trials with an adaptive choice of hypotheses, *Drug Information Journal*, **33**, 1423-1429.

Jennison, C. and Turnbull, B. W. (1999), *Group Sequential Methods with Applications to Clinical Trials*, Chapman & Hall.

Lan, K. K. G. and DeMets, D. L. (1983), Discrete sequential boundaries for clinical trials, *Biometrika*, **70**, 659-663.

Lan, K. K. G. and Trost, D. C. (1997), Estimation of parameters and sample size re-estimation. *Proceedings of Biopharmaceutical Section: American Statistical Association*, 48-51.

Lehmacher, W. and Wassmer, G. (1999), Adaptive sample size calculations in group sequential trials, *Biometrics* **55**, 1286-1290.

Little, R. J. A. and Rubin, D. B. (1987), *Statistical Analysis with Missing Data*, John Wiley & Sons.

Liu, Q. (2002), Personal Communication.

Liu, Q. and Chi, G. Y. H. (2001), On sample size and inference for two-stage adaptive designs, *Biometrics*, **57**, 172-177.

Liu, Q., Proschan, M. A. and Pledger, G. W. (2002), A unified theory of two-stage adaptive designs, *Journal of the American Statistical Association*, **97**, 1034-1041.

Marcus, R., Peritz, E. and Gabriel, K. R. (1976), On closed testing procedures with special reference to ordered analysis of variance, *Biometrika*, **63**, 655-660.

Morikawa, T. and Yoshida, M. (1995), A Useful Testing Strategy in Phase III Trials: Combined Test of Superiority and Test of Equivalence, *Journal of Biopharmaceutical Statistics*, **5**(5), 297-306.

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

Posch, M. and Bauer, P. (1999), Adaptive two stage designs and the conditional error function, *Biometrical Journal*, **6**, 689-696.

Posch, M., Bauer, P. and Brannath, W. (2003), Issues in designing flexible trials, *Statistics in Medicine*, **22**, 953-969.

Proschan, M. A. and Hunsberger, S. A. (1995), Designed extension of studies based on conditional power, *Biometrics*, **51**, 1315-1324.

Shun, Z. (2001), Sample size reestimation in clinical trials, *Drug Information Journal*, **35**, 1409-1422.

Snapinn, S. M. (2000), Noninferiority trials, *Current Controlled Trials in Cardiovascular Medicine*, **1**, 19-21.

Wang, S. J., Hung, H. M. J., Tsong, Y. and Cui, L. (1999), Group sequential test strategies for superiority and non-inferiority hypotheses in active controlled clinical trials, *Proceedings of the Biopharmaceutical Section of the American Statistical association*, 179-184.

Wang, S. J., Hung, H. M. J., Tsong, Y. and Cui, L. (2001), Group sequential test strategies for superiority and non-inferiority hypotheses in active controlled clinical trials, *Statistics in Medicine*, **20**, 1903-1912.

Wassmer, G. (1998), A comparison of two methods for adaptive interim analysis in clinical trials, *Biometrics*, **54**, 696-705.

Wittes, J. and Brittain, E. (1990), The role of internal pilot studies in increasing the efficiency of clinical trials (with comments), *Statistics in Medicine*, **9**, 65-72.

Zucker, D. M., Wittes, J. T., Schabenberger, O. and Brittain, E. (1999), Internal pilot studies II: comparison of various procedures, *Statistics in Medicine*, **18**, 3493-3509.