# TWO-STAGE WINNER DESIGNS FOR NON-INFERIORITY TRIALS WITH TWO TO THREE EXPERIMENTAL TREATMENTS AND AN ACTIVE CONTROL

**By PIN-WEN WANG**

**A Dissertation submitted to the**

**School of Public Health**

**University of Medicine and Dentistry of New Jersey**

**and the**

**Graduate School-New Brunswick**

**Rutgers, The State University of New Jersey**

**In partial fulfillment of the requirements**

**for the degree of**

**Doctor of Philosophy**

**UMDNJ-School of Public Health**

**Awarded jointly by these institutions and**

**Written under the direction of**

**Shou-En Lu, Ph.D. and Yong Lin, Ph.D.**

**And approved by**

_____

_____

_____

_____

**Piscataway/New Brunswick, New Jersey**

**May, 2013**

## ABSTRACT OF THE DISSERTATION

# TWO-STAGE WINNER DESIGNS FOR NON-INFERIORITY TRIALS WITH TWO TO THREE EXPERIMENTAL TREATMENTS AND AN ACTIVE CONTROL

**By PIN-WEN WANG**

**Dissertation Directors:**
**Shou-En Lu, Ph.D. and Yong Lin, Ph.D.**

In drug development, two-stage winner design can be cost-effective when the best treatment is to be determined from multiple experimental treatments. In this design, an interim analysis is devised to select the best treatment to avoid high cost, long-term trial conduction, and exposure to ineffective treatments. When an existing effective treatment is available, but new experimental treatments have advantages such as less cost, easier delivery, less invasive and fewer side effects, etc., including placebo in the trial may be considered unethical. It is desirable to conduct a non-inferiority trial that directly compares new treatments with the existing treatment (active control), and show that the new treatments are not less effective than the active control by a certain amount (so called non-inferiority margin).

In this dissertation, we extended the framework of the two-stage winner design of Shun et al. (2008) to conduct non-inferiority tests. Specifically, we considered designs of trials with two or three experimental treatments and an active control. Our methods include superiority hypotheses as a special case, but the hypothesis setting is more general than that studied by Shun et al.. We studied the distribution of test statistics, cut-off values, sample size and power calculations using exact distribution of the test statistics as well as using

normal approximations. Theoretical justifications and extensive numerical assessments were conducted to calculate the design parameters and evaluate the performance of our methods.

# Acknowledgements

I would like to express my deepest gratitude and appreciation to my dissertation advisors, Dr. Shou-En Lu and Dr. Yong Lin. This work would not exist without their supreme patience and intellectual generosity. During my years at school, I have known them as great statisticians and excellent educators. They have provided numerous valuable ideas, constant encouragement, and difficult long-distance consultation during my dissertation years. I am privileged to learn from their research experiences.

Dr. Lu has advised me since the day I entered UMDNJ's graduate program. During these years, she has always been a pillar that I can rely on throughout the ups and downs of my doctoral journey. Her insights were indispensable that eventually brought me to this exciting research topic. I appreciate her for the contribution of time, energy as well as numerous statistical ideas that were proven to be the key points of the dissertation. Furthermore, she taught me scientific writing and has edited a myriad of drafts that help set this dissertation in the right direction. She is a role model as a biostatistician, mentor and teacher. I still think fondly of my time as a research assistant for her.

Dr. Lin's superb advice and insights on mathematical subtlety and statistical computation were critical for tackling several theoretical hurdles and computational challenges for the research of this dissertation. Without Dr. Lin's many ideas, this dissertation would not be complete, not to mention that he had overcome the challenges and difficulties in reviewing and editing my countless drafts from a distance. Dr. Lin has set an example that I can follow as a fellow biostatistician.

I am grateful to Dr. Weichung Joe Shih on having assembled the talented and inspiring faculty, staff and students in the department. Dr. Shih has guided and inspired me for my entire Ph.D. study. He gave me several opportunities to work on a variety of research projects. His insights, perspectives and comments broadened my views and knowledge in

biostatistics research as well as helped developing my critical thinking capacity.

I am indebted to Dr. Gordon K.K. Lan for being on my committee, giving inspirations on my dissertation research, reviewing my dissertation, and providing great comments.

I would also like to thank the department's faculty members as they have helped me building a strong foundation in statistics. I am thankful and honored to be funded by Dr. Lu's research funding, which allowed me focusing on the research. I am also honored for receiving the SPH General Scholarship and I would like to offer my gracious thanks to the donors, the UMDNJ foundation, and the School of Public Health faculty.

I would like to offer heartfelt thanks to all my friends who have given me a hand during my PhD study. In particular, they include Yingshan You, Jin Hung, Po-Ting Lin, Ariel Chen, Ching-Jui Chang, Elisa Yang, Carol Chung, Xin Li and Linda Yau.

Lastly, I would like to thank my family for all their love and encouragement. My parents raised me to develop the love of mathematics and supported me in all my pursuits. For the company of my brother here in the US for five years and my husband who took care of my kids while I was traveling to take courses. My lovely daughters who took care of themselves, sacrificed their family time to be with me so I could focus on my study and work in these many years. I love them.

# Dedication

To my parents, my brother, Joyce, Jessica, Jacqueline and Hung-chih.

# Table of Contents

# List of Tables

# List of Figures

# Chapter 1

# Introduction

## 1.1 Background

With a large number of proven efficacious treatments, it is often unethical to use a placebo, a non-effective treatment control, or a very low dose of an active drug for patients with life-threatening or irreversible, life-altering conditions. The World Medical Association developed the Declaration of Helsinki [1] to provide the ethical principles for medical research involving human subjects. The Declaration states:

"The benefits, risks, burdens and effectiveness of a new intervention must be tested against those of the best current proven intervention, except in the following circumstances: the use of placebo, or no treatment, is acceptable in studies where no current proven intervention exists; or where for compelling and scientifically sound methodological reasons the use of placebo is necessary to determine the efficacy or safety of an intervention and the patients who receive placebo or no treatment will not be subject to any risk of serious or irreversible harm. Extreme care must be taken to avoid abuse of this option."

The International Conference on Harmonization Guidance on Choice of Control Group and Related Issues in Clinical Trials (ICH$E$10) [2] also states:

"In cases where an available treatment is known to prevent serious harm, such as death or irreversible morbidity in the study population, it is generally inappropriate to use a placebo control."

"In other situations, when there is no serious harm, it is generally considered ethical to ask patients to participate in a placebo-controlled trial, even if they may experience discomfort as result, provided the setting is noncoercive and patients are fully informed about available therapies and the consequences of delaying treatment."

If a placebo-controlled trial is considered to be unethical due to the availability of a treatment that is known to prevent serious harm, such as death or irreversible morbidity, it is desirable to conduct an active-controlled trial in which the experimental treatment is compared directly to a proven effective active control. Recently, the term "non-inferiority" has been increasingly used in clinical trials for this purpose. A therapeutic threshold known as the "non-inferiority margin", denoted by $\epsilon$, is needed to quantitatively define the clinical indifference. The non-inferiority margin is also closely related to the selection of the scale of the treatment effect. The pros and cons for a difference scale versus a ratio scale were extensively discussed in the literature, such as Brown and Day (2006), Hauschke and Hothorn (2007), and the references cited therein.

The use of adaptive designs based upon accrued data has become popular in clinical research and development because of its flexibility and efficiency in modifying the study design aspects of ongoing clinical trials. The adaptive designs allow to increase the sample size using the interim outcome to ensure study power or to stop the trial early without the burden of amending the protocol [3]. Even though the objectives for most of the adaptive design trials are testing superiority, it is possible to use an adaptive design for non-inferiority hypotheses. A seamless trial is a clinical trial design that combines multiple objectives that are traditionally addressed in separate trials into a single trial. An adaptive seamless design is a multi-stage design that provides several adaptive advantages such as stopping the trial early for efficacy/futility, sample size re-estimation or "drop the losers," etc. [4]. The primary objective of an adaptive seamless design is often to combine the phases of "dose selection" and "dose confirmation" into one trial. Such a design not only reduces the time between the learning and confirmatory phase by selecting and confirming the optimal dose in one study, but also combines the data from the learning phase and the confirmatory phase in the final analysis. Two-stage seamless adaptive designs are probably the most frequently used design because of their simplicity and because they combine two separate studies into one single study [5]. The two-stage winner trial is a special case of the two-stage seamless adaptive design, because it choose the "winner" treatment at the interim look. Lan et al. (2005) [6] proposed a simple way to use normal approximation on the test statistics to

evaluate the critical values of the two-stage winner deign. Shun et al. (2008) [7] extended the methodology to estimate the sample size using normal approximation. Practical strategies that consider both the multiplicity and adaptability related to the interim treatment selection are recommended.

This dissertation is motivated by the examples of LUCENTIS® studies (detailed below) that described the need of determining a more practical dosing regimen group using a two-stage seamless design. Since LUCENTIS® has been approved as an effective treatment for improving visual acuity for patients with age-related macular degeneration (AMD) and it has been on the market for other indications, it is more appropriate to compare the new dosing regimens to the approved dosing regimen (active control) than to the placebo directly. Because one needs only to demonstrate that the new dosing regimen is no less effective than the approved regimen, it is preferable to employ the non-inferiority test for this purpose. So far, to the best of our knowledge, there has been no discussion in the literature for a non-inferiority hypothesis using a two-stage winner design. In this thesis, we aimed to propose a statistical methodology for designing a two-stage winner design for testing non-inferiority hypotheses.

## 1.2 Motivating Examples

In the pivotal clinical studies of LUCENTIS® for patients with age-related macular degeneration (AMD) [8, 9, 10], it has been demonstrated that ongoing monthly dosing for a duration of 2 years is an effective therapy for improving visual acuity (VA). This monthly regimen required patients to go to a clinic to receive an injection of LUCENTIS® into the affected eyes. Serious eye infection, detached retina and other side effects have occurred with LUCENTIS® treatment. Additionally, such repetitive and invasive treatment can be impractical for elderly patients.

Therefore, a Phase 3*b* study [11] was conducted to explore an alternative regimen with less frequent dosing. From the results in all related studies, monthly dosing for the first 4 months, followed by an "as needed" dosing regimen was recommended. However, the

as needed plan did not provide a cogent guidance to physicians whether to adopt a less frequent regimen after the first 4 months, nor did it adequately address whether a physician could increase the recommended dose in less frequent injections and still achieve the same effect observed in the pivotal clinical studies.

Similar challenges were faced in the design of pivotal trials of LUCENTIS® for macular edema following retinal vein occlusion (RVO) [12]. The study design of one of the pivotal trials is shown in Figure 1.1. As shown in Figure 1.1, this study also included the "as needed" dosing plan after the recommended, mandatory injections for the first 6 months. Because LUCENTIS® had been approved as an effective treatment for both AMD and RVO, it was unethical to design a study to compare the high-dose less-frequent regimen of LUCENTIS® to a placebo control. The two-stage design winner design for non-inferiority hypotheses seem to be the most appropriate and practical design to test the high-dose less-frequent hypotheses. In this design, we could choose a "winner" dose regimen at interim look and perform a non-inferiority test in the final analysis to compare the treatment effect of the "winner" versus the original approved dose/regimen plan of LUCENTIS®. This example demonstrates the need for designing a two-stage winner trial for non-inferiority (NI) hypotheses that can adequately answer the following interested questions: 1) What should be an appropriate dose/regimen group after the mandatory dosing period? and 2) Whether the treatment effect of the "winner" dose/regimen is no worse than the treatment effect being seen in the pivotal trials within an acceptable range?

## 1.3 Research Objectives

The following are the objectives of this thesis:

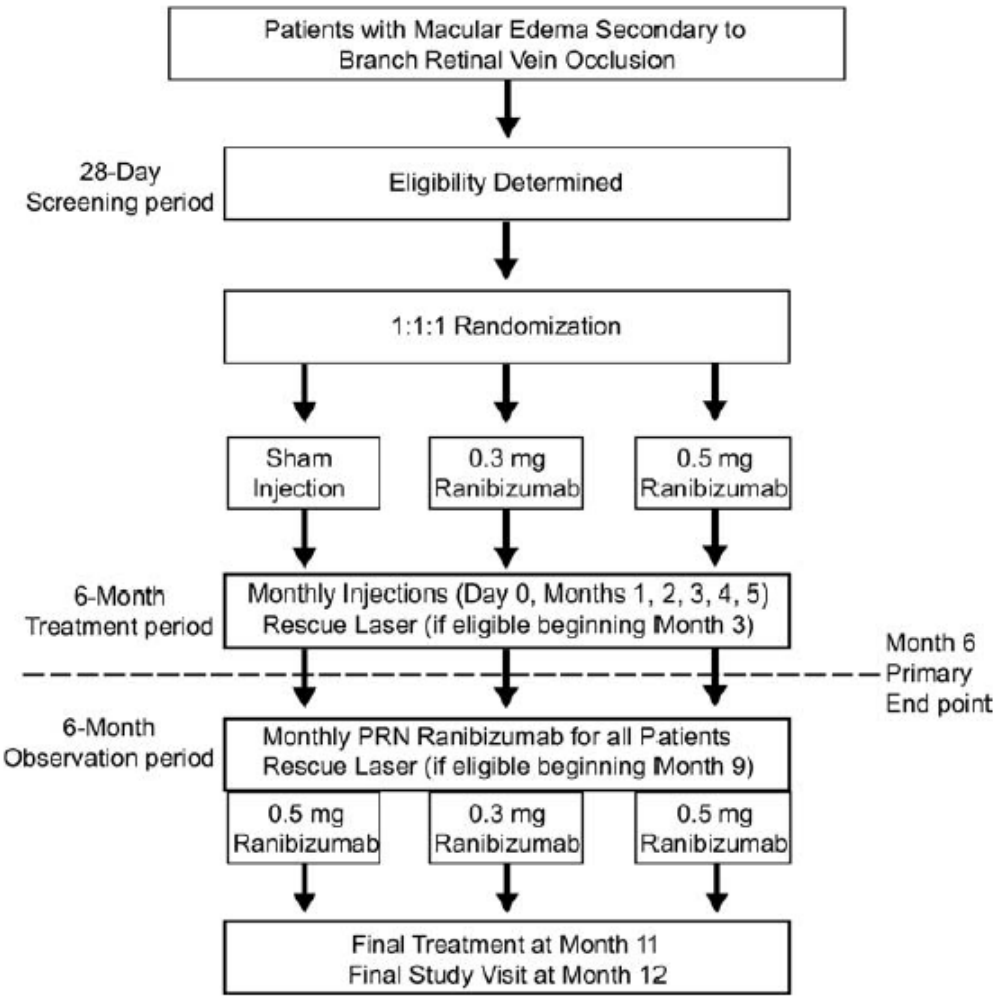
1. Propose a methodology to design a non-inferiority trial with a pre-specified non-inferiority margin under a two-stage winner design with two experimental treatment and one active control arms.

2. Propose a methodology to design a non-inferiority trial or a superiority trial under a two-stage winner design with three experimental treatments and one control arms.

## 1.4 Overview of the Remaining Chapters

This thesis is organized as follows. In Chapter 2, we reviewed the statistical methodologies of non-inferiority trials followed by the review of the two-stage winner design using normal approximation. In Chapter 3, we proposed a methodology to conduct non-inferiority trials with a pre-specified non-inferiority margin under the two-stage winner design frame work. In Chapter 4, we extend the work in Chapter 3 to conduct a two-stage winner design with three experimental treatment and one control arms. Both superiority and non-inferiority hypotheses are considered. In Chapter 5, we conclude the methodologies and discuss some possible future work.

Figure 1.1: Example of study design- Ranibizumab for macular edema following branch retinal vein occlusion: six-month primary end point results of a phase 3 study, excerpted from Campochiaro et al. (2010)



# Chapter 2

# Literature Review

The main reasons for using a non-inferiority trial are usually ethical. Specifically, we would want to avoid exposing patients to an unsafe or ineffective treatment if a proven effective treatment is already available [2]. Therefore, it may be more appealing to evaluate whether a new experimental treatment is equal or not inferior to an existing efficacious therapy, thus testing the usual hypothesis of equality in a conventional superiority trial can be inappropriate. Blackwelder (1982) [13] proposed a new idea of "Proving the Null Hypothesis" that it focuses on testing whether a new or experimental treatment is as effective as control/standard therapy (equivalence). This idea differs from the conventional thinking of proving the superiority of a new or experimental treatment over the control/standard therapy. The concept of proving "non-inferiority" was further officially defined in the document "Points to consider on switching between superiority and non-inferiority" [14] by the Committee for Proprietary Medicinal Products (CPMP), an European Union regulatory agency, as well as in the discussion within the Food and Drug Administration (FDA).

Table 2.1 summarizes the differences between a non-inferiority test (with a fixed margin) versus the conventional superiority test for normally distributed data. In this table, $\epsilon$ is a sufficiently small number to specify the equivalency of therapies for a non-inferiority hypothesis. It is called the *non-inferiority margin*, that makes the denominator of the sample size formula larger for a non-inferiority test. Thus, the required sample size for a non-inferiority trial is in general smaller than that for a conventional superiority trial. In late 1990s, the distinction between equivalence and non-inferiority was made officially in the document "Points to consider on switching between superiority and non-inferiority" by the Committee for Proprietary Medicinal Products (CPMP) [14], an European Union regulatory agency, as well as in discussions within the Food and Drug Administration (FDA).

In this thesis, we proposed the methodology using the two-stage winner designs for non-inferiority hypotheses. Therefore, we will review the literatures related to the non-inferiority hypotheses/tests and related to the two-stage design in the following sections.

## 2.1 Non-inferiority Hypotheses and Tests

Conceptually, non-inferiority trial provides two comparisons:

1. A direct comparison between the experimental treatment with the active control.
2. An indirect comparison between the experimental treatment and the placebo, with an estimated treatment effect of active control from historical trials.

Both comparisons involved an active control arm, and the validity of a non-inferiority trial strongly relies on the accuracy of its estimated treatment effect. Two major objectives are often proposed in non-inferiority trials [15]:

- To test whether an experimental treatment is efficacious in the sense that it would have been more effective than a placebo had a placebo been in the trial.
- To infer that an experimental treatment retains a proportion of an active control's effect (as relative to the placebo).

Before elaborating these objectives, we first introduce some notations. Assume that the efficacy of the treatment is measured on some scale such that a larger value represents better efficacy. Denote the true effect of the experimental treatment (E) relative to active control (C) by $\beta_{EC}$, and denote the true effect of the active control (C) relative to placebo (P) by $\beta_{CP}$. Thus, the true effect of the experimental treatment relative to placebo can be represented by: $\beta_{EC} + \beta_{CP}$. In order to test the efficacy of an experimental treatment relative to a placebo that not present in current trial (related to the first objective), we can formulate the statistical hypotheses as:

$$H_0 : \beta_{EC} + \beta_{CP} \leq 0 \text{ versus } H_1 : \beta_{EC} + \beta_{CP} > 0.$$

Equivalently,

$$H_0 : \beta_{EC} \leq -\beta_{CP} \text{ versus } H_1 : \beta_{EC} > -\beta_{CP}.$$

In this context, the non-inferiority margin $\epsilon = \beta_{CP}$ is the effect of the active control over the placebo. In order to infer the proportion of retention of an active control effect for an experimental treatment (related to the second objective), we can formulate the relevant statistical hypotheses as:

$$K_0 : \beta_{EC} + \beta_{CP} \leq u\beta_{CP} \text{ versus } K_1 : \beta_{EC} + \beta_{CP} > u\beta_{CP},$$

where $u$ is the proportion of the control effect to be retained ($0 \leq u \leq 1$). Equivalently,

$$K_0 : \beta_{EC} \leq -(1-u)\beta_{CP} \text{ versus } K_1 : \beta_{EC} > -(1-u)\beta_{CP}.$$

Thus, the non-inferiority margin with the effect retention can be presented as $\epsilon = (1-u)\beta_{CP}$. Because $(1-u)\beta_{CP} < \beta_{CP}$, the margin needs to demonstrate the effect retention is narrower than the margin to demonstrate efficacy of the experimental treatment. Therefore, the conclusion of effect retention can infer the efficacy of the experimental treatment. For some clinical endpoints, such as mortality, the prior objective of simply showing the experimental treatment being superior to a placebo is deemed not enough. As a result, the second objective is often required in a regulatory application, especially for the active control treatments that have already demonstrated large effect. However, determining the percent retention can be as difficult as the determination of a fixed margin. The logical inconsistencies associated with the use of hypotheses $K_0$ had brought some attentions and discussed elsewhere [16]. In this dissertation, we will assume the non-inferiority margin is known and is pre-determined.

Table 2.1: Comparison between the conventional and non-inferiority hypotheses for normally distributed data, formulas[a] adopted from Blackwelder(1982)

| | Conventional hypothesis test | NI test with fixed margin $\epsilon$ |
|---|---|---|
| Null hypotheses | $H_0 : \mu_E - \mu_C \leq 0$ | $H_0' : \mu_E - \mu_C \leq -\epsilon$ |
| Alternative hypotheses | $H_1 : \mu_E - \mu_C > 0$ | $H_1' : \mu_E - \mu_C > -\epsilon$ |
| Test statistic | $Z = \frac{\bar{X}_E - \bar{X}_C}{SE}$ | $Z' = \frac{\bar{X}_E - \bar{X}_C + \epsilon}{SE}$ |
| Standard error (SE) | $\sqrt{\frac{{\sigma_C}^2}{n_C} + \frac{{\sigma_E}^2}{n_E}}$ | $\sqrt{\frac{{\sigma_S}^2}{n_C} + \frac{{\sigma_E}^2}{n_E}}$ |
| 100(1-$\alpha$)%confidence interval | $(\bar{X}_E - \bar{X}_C - Z_{1-\alpha} * SE, \infty)$ | $(\bar{X}_E - \bar{X}_C - Z_{1-\alpha} * SE, \infty)$ |
| Sample size in each of two groups* | $\frac{(Z_{1-\alpha}+Z_{1-\beta})^2[{\sigma_C}^2+{\sigma_E}^2]}{(\mu_E-\mu_C)^2}$ | $\frac{(Z_{1-\alpha}+Z_{1-\beta})^2[{\sigma_C}^2+{\sigma_E}^2]}{(\mu_E-\mu_C+\epsilon)^2}$ |
| Sample size calculation is condition on | $\mu_C < \mu_E$ | $\mu_E - \mu_C > -\epsilon$ |

[a] $\epsilon$ is the non-inferiority margin. $\mu_C$ and $\mu_E$ are the true mean for control and experimental treatment, respectively.
${\sigma_C}^2$ and ${\sigma_E}^2$ are the true variance for control and experimental treatment, respectively.
$\bar{X}_C$ and $\bar{X}_E$ are the corresponding observed sample means, and $n_C$ and $n_E$ are the numbers of patients in the two groups.
$Z_{1-\alpha}$ and $Z_{1-\beta}$ are upper percentage points of the standard normal distribution.
* Assume equal sample size in each group.

There are several challenges in designing a non-inferiority trial and forming hypotheses. One of the major challenges is on the determination of a non-inferiority margin, which will be discussed in details in the next section. The other challenges are mostly related to the following two important assumptions: assay sensitivity and constancy assumption [17]. The former refers to the ability of a clinical trial that can demonstrate a difference between treatments if there is truly a difference. The latter refers to the similarity between a non-inferiority trial and historical studies with respect to all important design and conduct features that might influence the effect size of the active control. Simply put, it requires that the effect size of the active control in the non-inferiority trial under study is similar to that in the historical trials. There are a number of factors that can undermine the assay sensitivities, such as poor compliance with treatment, poor selection of study population, and biased endpoint assessment, etc. [2]. The constancy assumption can be undermined by factors such as different patient enrollment criteria, dose regiment of the control, concomitant medication, and inconsistent results in the historical placebo-controlled trials, etc. In order to have a successful and credible non-inferiority trial, we must attempt to avoid these factors to every possible extent. The focus of this thesis is not on validating these assumptions. Therefore, we will assume the assumptions are met in our discussions.

### 2.1.1 Rationale of Non-inferiority Trials

Consider a non-inferiority trial with two experimental treatment arms: one experimental treatment and one active control. One of the goals for a non-inferiority trial is to show that an experimental treatment is no worse than a control within an acceptable range. We consider the following hypotheses:

$$H_0 : \beta_{EC} \leq -\epsilon \text{ ( E is inferior to the control by } \epsilon \text{ or more)},$$
$$H_1 : \beta_{EC} > -\epsilon \text{ ( E is not inferior to the control by more than } \epsilon).$$

where $\epsilon$ represents the entire effect of the active control assumed to be present in the non-inferiority study (relative to the placebo).

The outcome of a non-inferiority trial is usually assessed by a two-sided 95% confidence interval (or one-sided 97.5% interval), for evaluating the true differences between an experimental treatment and an active control [18]. The outcome usually are summarized in the following two quantities:

- The point estimate of the estimated effect (i.e. the observed difference between E and C).
- The lower limit of the confidence interval of the estimated effect.

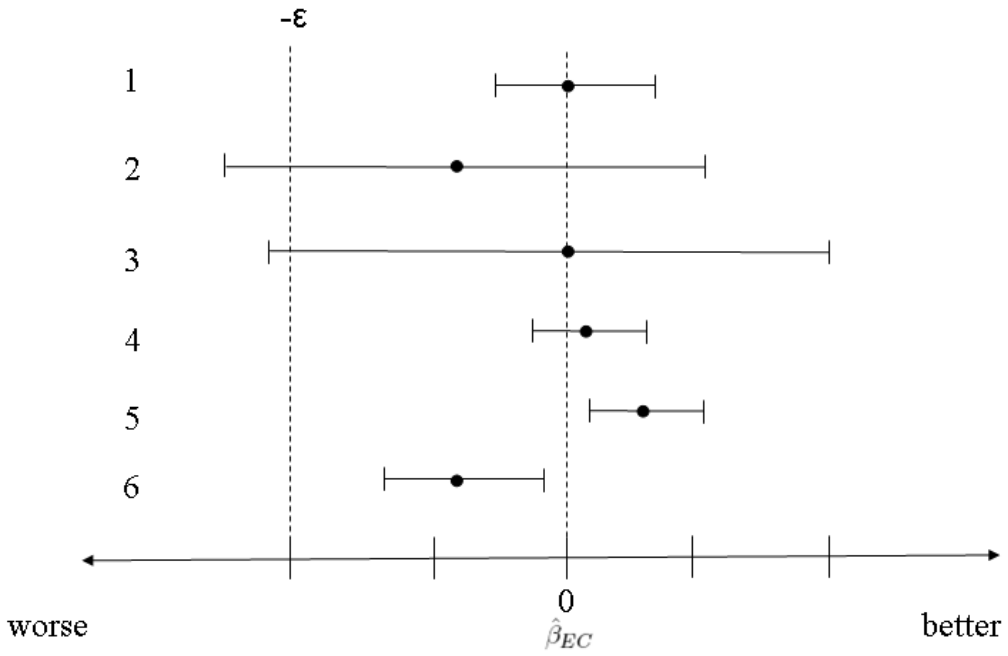
The former represents the best estimate of the true difference between the two treatments, and the later represents a lower bound that is usually interpreted as the degree of inferiority to the reference that can be excluded based on the data observed (the chance that the true difference is worse than this bound is acceptably small). If the E and C are equally efficacious, there is an equal chance that the point estimate will be positive or negative regardless of the sample size. For this reason, the point estimate alone is not sufficient as an estimate of the relative efficacy. However, in the situation of true equality, the lower bound of confidence interval would expect to move closer to 0 as the sample size increased [17]. Figure 2.1 shows various scenarios of the estimated treatment effects $\hat{\beta}_{EC}$ and the corresponding two-sided 95% confidence intervals (CI) on a difference scale $E - C$)in non-inferiority trials. The non-inferiority can be established if the lower bound of the two-sided confidence interval for $\hat{\beta}_{EC}$ is greater than $-\epsilon$. The interpretation of the results for the scenarios are:

1. Point estimate of $\hat{\beta}_{EC}$ is 0, suggesting equal effect; lower bound of the 95% CI for $\hat{\beta}_{EC} > -\epsilon$; non-inferiority is demonstrated.
2. Point estimate of $\hat{\beta}_{EC}$ favors C; lower bound of the 95% CI for $\hat{\beta}_{EC} < -\epsilon$; non-inferiority is not demonstrated.
3. Point estimate of $\hat{\beta}_{EC}$ is 0, suggesting equal effect; but the lower bound of the 95% CI of $\hat{\beta}_{EC} < -\epsilon$, so that non-inferiority is not demonstrated.
4. Point estimate favors E; non-inferiority is demonstrated, but superiority is not demonstrated.

5. Point estimate favors E; superiority and non-inferiority are demonstrated.

6. Point estimate of $\hat{\beta}_{EC}$ favors C and C is statistically significantly superiority than E. However, the lower bound of the 95% for $\hat{\beta}_{EC} > -\epsilon$, so that non-inferiority is demonstrated with the given non-inferiority margin $\epsilon$ (This outcome would be unusual and could present interpretive problems).

Figure 2.1: Results of the estimated treatment effect of $\hat{\beta}_{EC}$ and 95% confidence interval(CI)




As mentioned above, the major challenge in designing, conducting and interpreting the result of a non-inferiority trials is the determination of a non-inferiority margin $\epsilon$, which is not measured in the non-inferiority study due to the lack of concurrent placebo arm [19]. There are abundant discussions in the literature regarding how to determine this quantity (e.g., Wiens (2002), Hung et al. (2003), Chow and Shao (2006), Hung et al. (2003,2005,2007,2009), D'Agostino Sr. et al. (2003), Ng (2008) and the references cited therein). The determination of a non-inferiority margin $\epsilon$ depends on several important factors [17]:

- The treatment effect estimated from the historical trials with the active control.
- The assessment of the likelihood that the effect of the active control is similar to the its effect in the historical trials (constancy assumption).
- The assessment of the quality of the non-inferiority trial (therefore, the size of the $\epsilon$ cannot be entirely specified until the non-inferiority study is complete).

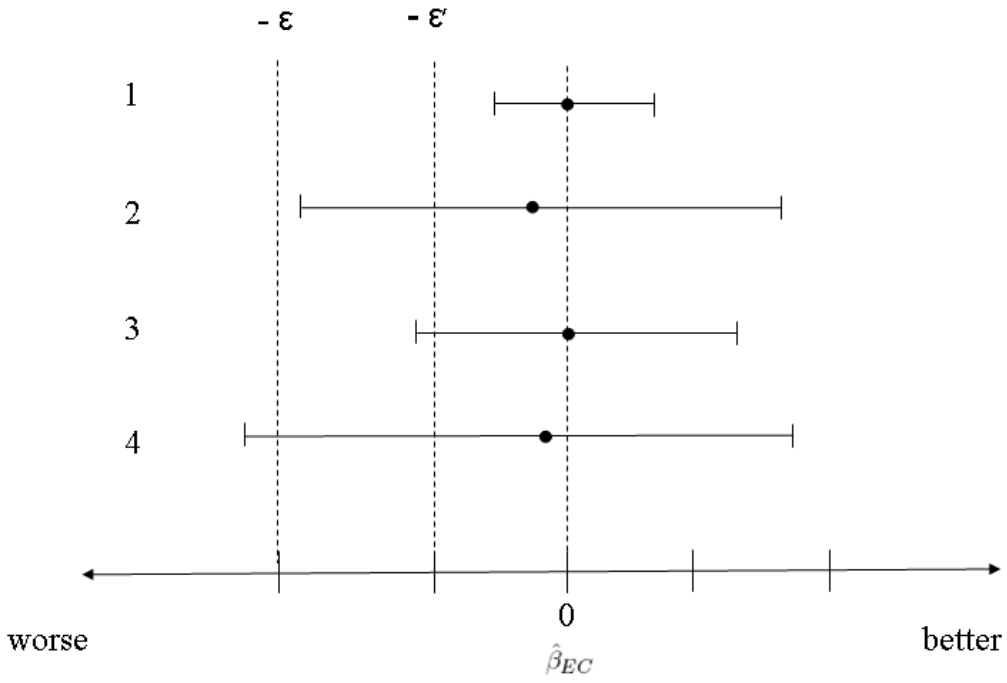
The validity of any conclusion from a non-inferiority trial relies on the choice of $\epsilon$. In practice, we usually choose a smaller margin $\epsilon'$ that reflects the largest clinically acceptable difference (degrees of inferiority) of the experimental treatment compare to the active control. To illustrate this margin in relation to the $\epsilon$, Figure 2.2 shows various scenarios on the estimated treatment effect of $\hat{\beta}_{EC}$ and the corresponding two-sided 95% CI with the assumed non-inferiority margins $\epsilon$ and $\epsilon'(< \epsilon)$. The interpretation of the results are as follows:

1. Point estimate of $\hat{\beta}_{EC}$ is 0; lower bound of the 95% for $\hat{\beta}_{EC} > -\epsilon'$ , indicating experimental treatment is effective (non-inferiority is demonstrated).
2. Point estimate of $\hat{\beta}_{EC}$ favors C; lower bound of the 95% for $\hat{\beta}_{EC}$ is between the $-\epsilon$ and $-\epsilon'$, indicating there is a positive but unacceptable loss of the control effect.
3. Point estimate of $\hat{\beta}_{EC}$ is 0; lower bound of the 95% for $\hat{\beta}_{EC}$ is between the $-\epsilon$ and $-\epsilon'$. This could lead to the effectiveness conclusion.
4. Point estimate of $\hat{\beta}_{EC}$ favors C and lower bound of the 95% CI $< -\epsilon$, suggesting there is no evidence of effectiveness for the experimental treatment.

The $\epsilon'$ can never be greater than $\epsilon$ and is determined from the clinical judgement. Due to this reason, there may be a greater flexibility in interpreting the 95% lower bound for the effect of $\hat{\beta}_{EC}$ that is slightly smaller than $-\epsilon'$, but the lower bound is still larger than $-\epsilon$, as shown in Figure 2.2 scenario 3.

To test non-inferiority hypotheses under the frequentist framework, there are usually two classes of statistical methods. These methods are often referred as the fixed margin

Figure 2.2: Results of the estimated treatment effect of $\hat{\beta}_{EC}$ and 95% confidence interval(CI) from non-inferiority trials in relation with $\epsilon$ and $\epsilon'$, excerpted from [17]




approach [17, 15, 20] and the synthesis approach [21, 17, 15, 20, 22, 23]. The goal of the fixed margin approach is to demonstrate that the effect of an experimental treatment is no worse than the effect of an active control by more than a pre-specified non-inferiority margin. The determination of a non-inferiority margin depends not only from the statistical evidence, but also depends on the clinical judgement. In contrast, the synthesis method does not require to pre-specified a non-inferiority margin but it combines the point effect estimate and variance from the non-inferiority trial with those estimates obtained from the historical trials. Such a combination bridges the data from historical trials with the data from non-inferiority trial into one test statistic. Therefore, we could make inferences on the experimental treatment effect relative to the placebo that is not present in the current trial. In summary, the fixed margin approach highlights the multi-dimensional statistical and clinical considerations that are essential in determining the non-inferiority margin [20]. The process of the non-inferiority margin determination captures the statistical uncertainty by using the standard error of the estimated relative effect of the active control from historical data. On the contrary, the synthesis approach does not generate a fixed margin nor utilizing

any clinical margins on making statistical inferences. It incorporates the standard errors (from both the historical and non-inferiority trial) directly into the construction of a test statistic for statistical inferences. In the next section, we provide detailed review for each of these methods. First, we will formulate the hypotheses of interest in this thesis:

$$H_0 : \beta_{EC} + \beta_{CP} \leq 0,$$

$$H_1 : \beta_{EC} + \beta_{CP} > 0.$$

### 2.1.2 Fixed Margin Approach

To use the fixed margin approach, we first need to determine a non-inferiority margin based on clinical and statistical evidence. The determination of a non-inferiority margin usually involves two steps: Step 1: To obtain an effect estimate of the selected active control from the data in the historical placebo-controlled trials. Fixed-effect or random-effect meta-analytic approaches are the most commonly used [24, 25]. Step 2: To explore the extent of discounting, i.e., assess whether the effect of the active control can be presumed to be present in the new study (constancy assumption) or need to be adjusted based on the difference between the current non-inferiority trial and historical trials.

Various approaches have been used to determine the non-inferiority margin and to analyze an non-inferiority study. Let,

$$\hat{\beta}_{CP} \text{ and } \hat{V}_{CP}$$

denote the estimates of the treatment effect and variance for the active control relative to the placebo, respectively, while

$$\hat{\beta}_{EC} \text{ and } \hat{V}_{EC}$$

denote the estimates of the treatment effect and variance for the experimental treatment relative to the active control, respectively. Note that $\hat{\beta}_{CP}$ and $\hat{V}_{CP}$ are obtained from the historical trials, and $\hat{\beta}_{EC}$ and $\hat{V}_{EC}$ are obtained from the non-inferiority trial.

**95% − 95% method**

The two 95% confidence interval method is one of the most readily understood approach related to the fixed margin approach. This method was discussed in the FDA memorandum regarding the evaluation of thrombolytic therapies [26]. Assume that there is a set of historical trials comparing the control treatment with the placebo that have been pooled in some type of meta-analysis, and there is a new non-inferiority trial comparing the experimental treatment with the active control treatment. In most practice, a non-inferiority margin is usually determined based on historical placebo-controlled trials. According to the ICH E-10 guideline [2], a non-inferiority "margin cannot be greater than the smallest effect size that the active drug would be reliably expected to have compared with placebo in the setting of the planned trial." As a result, the first 95% refers to the confidence interval used to choose the effect size of the active control from the historical data, and the second 95% refers to the confidence interval used to reject the null hypothesis in the non-inferiority study. In order to ensure this, $\epsilon$ is often chosen based on the lower limit of the 95% confidence interval for the active control relative to placebo, specifically, $\epsilon = \hat{\beta}_{CP} - 1.96\sqrt{\hat{V}_{CP}}$. If the confidence interval for the effect of the experimental treatment relative to the active control is entirely above this margin, the experimental treatment can be declared to be non-inferior to the active control. In this context, the non-inferiority of an experimental treatment can be concluded if the lower bound of the two-sided confidence interval for the $\hat{\beta}_{EC}$ is greater than $-\epsilon$, i.e.,

$$\hat{\beta}_{EC} - 1.96\sqrt{\hat{V}_{EC}} > -\epsilon \left( \text{where } \epsilon = \hat{\beta}_{CP} - 1.96\sqrt{\hat{V}_{CP}} \right).$$

Equivalently,

$$\hat{\beta}_{EC} - 1.96\sqrt{\hat{V}_{EC}} > -\left( \hat{\beta}_{CP} - 1.96\sqrt{\hat{V}_{CP}} \right).$$

or

$$\frac{\hat{\beta}_{EC} + \hat{\beta}_{CP}}{\sqrt{\hat{V}_{EC}} + \sqrt{\hat{V}_{CP}}} > 1.96. \tag{2.1}$$

**95% − 95% method with discounting effect**

Another variation of the traditional 95% − 95% method is to employ a discounting factor

to the lower limit of the 95% CI for the relative effect of the active control. When there are factors that undermine the constancy assumption or when the truth of the assumption cannot be verified, it is necessary to employ a discounting factor, denoted by $u$, to the estimate of lower limit of the 95% CI for the effect of active control relative to placebo. Specifically, $\epsilon = (1-u)(\hat{\beta}_{CP} - 1.96\sqrt{\hat{V}_{CP}})$, where $0 \leq u \leq 1$. This is similar to applying the preservation of a fraction of the active control's effect to the non-inferiority margin. Once the margin is set prior to the conduct the non-inferiority trial, the margin can not be changed and will be viewed as a fixed constant that defines the statistical hypothesis. According to this approach, the non-inferiority can be established if $\hat{\beta}_{EC} - 1.96\sqrt{\hat{V}_{EC}} > -\epsilon$, where $\epsilon = (1-u)(\hat{\beta}_{CP} - 1.96\sqrt{\hat{V}_{CP}})$. Equivalently,

$$\frac{\hat{\beta}_{EC} + (1-u)\hat{\beta}_{CP}}{\sqrt{\hat{V}_{EC}} + \sqrt{(1-u)^2\hat{V}_{CP}}} > 1.96. \tag{2.2}$$

Once the statistical margin is being determined, we need to use other factors such as clinical significance, easiness of treatment administration, or cost, and other external information to determine a non-inferiority margin for the trial. This clinical margin can not be larger than the statistical margin. In Chapter 3 and Chapter 4, we will use the fixed margin approach assuming a pre-determined non-inferiority margin being determined from all available aspects.

### 2.1.3 Synthesis Approach

In contrast to the fixed margin approach, a non-inferiority margin is not required in the synthesis approach. The synthesis approach combines the statistical evidence from the historical trials with the one from the non-inferiority trial. One basic assumption supporting this method is that the control effect estimated from the historical trials can be carried over to the new non-inferiority trial. Once this assumption can be verified, we can then combine the effect estimate of the experimental treatment relative to the active control from the non-inferiority trial with the effect estimate of the active control to placebo from historical trials. Similar to the fixed-margin approach, there are different versions of the synthesis approach.

**Standard test without discounting**

Assume the assay sensitivity and constancy assumption hold. We also assume the estimated relative effect of the experimental treatment (from the current non-inferiority trial) and the estimated relative effect of the active control (from the historical trials) are independently and normally distributed. Thus, we could simply take the sum of the estimated effect of experimental treatment relative to active control, with the estimated effect of active control relative to placebo, to obtain a test statistic. The test statistic is approximately normally distributed with variance equal to the sum of the two variances of the relative effect estimates. The efficacy of the experimental treatment can be concluded if:

$$\frac{\hat{\beta}_{EC} + \hat{\beta}_{CP}}{\sqrt{\hat{V}_{EC} + \hat{V}_{CP}}} > 1.96. \tag{2.3}$$

**Preserve a fraction of the active control's effect(discounting effect)**

In the situation when the constancy assumption is questionable, we can apply a discounting factor to the estimated relative effect of the active control. Therefore, we reject the null hypothesis of $\beta_{EC} + (1-u)\beta_{CP} \leq 0$ in favor of $\beta_{EC} + (1-u)\beta_{CP} > 0$ to conclude the non-inferiority, where $0 \leq u \leq 1$. This can also be viewed as preserving a fraction of the active control's effect. Since these effect estimates are independent, the non-inferiority can be concluded if:

$$\frac{\hat{\beta}_{EC} + (1-u)\hat{\beta}_{CP}}{\sqrt{\hat{V}_{EC} + (1-u)^2\hat{V}_{CP}}} > 1.96. \tag{2.4}$$

**Summary of the Approaches**

The four test statistics described earlier are summarized in Table 2.2. We can see that the only differences between Fixed-margin and Synthesis approaches are on the denominator. Since

$$\sqrt{\hat{V}_{EC} + \hat{V}_{CP}} \leq \sqrt{\hat{V}_{EC}} + \sqrt{\hat{V}_{CP}},$$

| | Without discounting factor | With discounting factor |
|---|---|---|
| Fixed-margin approach | $\frac{\hat{\beta}_{EC}+\hat{\beta}_{CP}}{\sqrt{\hat{V}_{EC}}+\sqrt{\hat{V}_{CP}}}$ (2.1) | $\frac{\hat{\beta}_{EC}+(1-u)\hat{\beta}_{CP}}{\sqrt{\hat{V}_{EC}}+\sqrt{(1-u)^2\hat{V}_{CP}}}$ (2.2) |
| Synthesis approach | $\frac{\hat{\beta}_{EC}+\hat{\beta}_{CP}}{\sqrt{\hat{V}_{EC}+\hat{V}_{CP}}}$ (2.3) | $\frac{\hat{\beta}_{EC}+(1-u)\hat{\beta}_{CP}}{\sqrt{\hat{V}_{EC}+(1-u)^2\hat{V}_{CP}}}$ (2.4) |

Table 2.2: Excerpted from Snapinn(2004) [21]

and

$$\sqrt{\hat{V}_{EC}+(1-u)^2\hat{V}_{CP}} \leq \sqrt{\hat{V}_{EC}}+\sqrt{(1-u)^2\hat{V}_{CP}},$$

the fixed-margin approach is relatively more conservative. It is important to recognize that the synthesis approach makes use of the variabilities from both the non-inferiority trial and historical trials and yields one confidence interval in the statistical inference. Thus, the validity of this method strongly relies on the constancy assumption.

### 2.1.4 Unified Approach:

While various challenges associated with the design and analysis of the non-inferiority trials are known, many methodologies have been proposed, and it remains unclear what the most appropriate approach is. As the validity of a non-inferiority trial depends on two key assumptions: assay sensitivity and constancy, which sometimes can not be verified, Snapinn and Jiang (2008) [27] proposed an unified approach as a tool to control the type I error rate when there is a departure from the assumptions. This method can also explore the impact of variance inflation on the power. Using the unified approach $U(u,v)$, non-inferiority can be established if

$$U(u,v) > 1.96, \text{ where } U(u,v) = \frac{\hat{\beta}_{EC}+(1-u)\hat{\beta}_{CP}}{\sqrt{\hat{V}_{EC}+(1-u)^2\hat{V}_{CP}+2v(1-u)\sqrt{\hat{V}_{EC}\hat{V}_{CP}}}}. \quad (2.5)$$

In Equation 2.5, $v$ is a variance inflation factor. And $u$ is a discounting factor that have been discussed in the previous section. Note that $U(u,0)$ corresponds to the synthesis approach; $U(u,1)$ corresponds to the fixed-margin approach; and $U(1,v)$ corresponds to a test for the

superiority of the experimental treatment to the active control. In Chapter 4, we will use the test statistic of the unified approach in a two-stage winner design for non-inferiority hypotheses.

## 2.2 Existing Two-Stage Designs for Comparative Clinical Trials

The concept of treatment selection in two-stage design was proposed by Whitehead [28], Thall et al. [29, 30] and Schaid et al. [31] for binary and survival endpoints. Recently, the unconditional and conditional procedure were considered by Stallard et al. and Sampson et al. [32, 33, 34] to select one-winner for normally distributed endpoints. Friede et al. [35] gave a comparison for adaptive treatment selection. Liu et al. [36], Koenig et al. [37], and Li et al. [38, 39] also considered the procedure to select more than one winners. Shun et al.(2008) proposed to use a normal approximation method to design a two-stage winner trial. At interim look, we select one of the two treatment arms with better treatment effect to continue with the control arm to the final analysis. In this thesis, we will extend the work from Shun et al. (2008) and will have detail review of their paper in the following section.

### 2.2.1 Review of Shun, Lan and Soo's Procedure

***Design, Test Statistic and Distribution:***

Suppose there is a study consists of 3 arms, one control and two treatment arms. Each arm consists of a sample size of n. One interim analysis at the information time $\tau$ ($= n_1/n$, where $n_1 \leq n$ is the interim sample size per group) is planned. Let $\{X_i^{(j)} | i = 1, ..., n_1\}$ and $\{Y_i^{(j)} | i = 1, ..., n\}$, $j = 0, 1, 2$, represent the continuous measurement at interim and final analysis, respectively. We use $j = 0$ to represent the control group, and $j = 1, 2$ to represent the two treatment groups. Moreover, we assume that $X_i^{(j)}$ are identically independently distributed with $N(\nu_j^X, \sigma_X^2)$ and $Y_i^{(j)}$ are identically independently distributed with $N(\mu_j^Y, \sigma_Y^2)$. We also assume that $corr(X_i^{(j)}, Y_i^{(j)}) = \rho$ for all $i = 1, ..., n_1$ and $j = 0, 1, 2$. In the case where the same endpoints are used at the interim and final analysis, $\rho$ will be 1. Let the interim sample means for each arm be $\bar{X}_{n_1}^{(0)}, \bar{X}_{n_1}^{(1)}, \bar{X}_{n_1}^{(2)}$, where $\bar{X}_{n_1}^{(j)} = (1/n_1) \sum_{i=1}^{n_1} X_i^{(j)}$, and

let the final sample means for each arm be $\bar{Y}_n^{(0)}, \bar{Y}_n^{(1)}, \bar{Y}_n^{(2)}$, where, $\bar{Y}_n^{(j)} = (1/n)\sum_{i=1}^{n} Y_i^{(j)}$. At the interim analysis, we keep only the treatment with greater sample mean and the control arm to the end of the study, assuming greater value is better.

Let $\Delta_j = \mu_j^Y - \mu_0^Y$ be the unknown treatment effect of the $j$'s treatment group compared to the control, $j = 1$ or 2. We consider the hypotheses:

$$H_0 : \Delta_1 = \Delta_2 = 0 \text{ versus } H_1 : \Delta_1 > 0 \text{ or } \Delta_2 > 0.$$

Assume the target treatment effect of the trial is $\Delta_j = \delta_j$, and let the test statistic be

$$Z_n^{(j)} = \sqrt{\frac{n}{2\sigma_Y^2}}(\bar{Y}_n^{(j)} - \bar{Y}_n^{(0)}), j = 1 \text{ or } 2.$$

Under $H_0$, $Z_n^{(j)} \sim N(0,1)$. Let

$$V_{n_1} = \sqrt{\frac{n_1}{2\sigma_X^2}}(\bar{X}_{n_1}^{(1)} - \bar{X}_{n_1}^{(2)}),$$

then the covariances between the $Z_n^{(j)}$ $(j = 1, 2)$ and $V_{n_1}$ are

$$\eta = cov(Z_n^{(1)}, V_{n_1}) = \frac{\sqrt{\tau}}{2}\rho = -cov(Z_n^{(2)}, V_{n_1}).$$

Under $H_0$, $V_{n_1} \sim N(0,1)$, and under $H_1$, $V_{n_1} \sim N(\lambda, 1)$, where $\lambda = \sqrt{\frac{n_1}{2\sigma_X^2}}(\nu_1^X - \nu_2^X) \neq 0$. For convenience, authors assumed that $\lambda > 0$.

The final test statistic of the study with an interim treatment selection is then

$$W = \begin{cases} Z_n^{(1)}, & \text{if } V_{n_1} > 0, \\ Z_n^{(2)}, & \text{if } V_{n_1} < 0. \end{cases}$$

We start with a general case under $H_1$ where $\delta_1 \neq \delta_2$ and $\lambda \neq 0$. Let $p = Pr(V_{n1} > 0)$ be the probability of selecting treatment 1 and $q = Pr(V_{n1} < 0) = 1 - p$. The distribution of

W is

$$\begin{aligned} F_W(w) &= Pr(W < w) \\ &= p\left[\frac{1}{p}\int_{-\infty}^{w-w_1} \Phi(k_0 + kz)\phi(z)dz\right] + q\left[\frac{1}{q}\int_{-\infty}^{w-w_2} \Phi(-k_0 + kz)\phi(z)dz\right] \\ &= pF_1(w - w_1) + qF_2(w - w_2), \end{aligned}$$

where

$$w_1 = \sqrt{\frac{n}{2\sigma_Y^2}}\delta_1, w_2 = \sqrt{\frac{n}{2\sigma_Y^2}}\delta_2,$$

$$k_0 = \frac{\lambda}{\sqrt{1-\eta^2}}, k = \frac{\eta}{\sqrt{1-\eta^2}},$$

and

$$F_j(w) = \int_{-\infty}^{w} f_j(t)dt, \text{ for } j = 1, 2,$$

with density functions

$$f_1(w) = \frac{1}{p}\Phi(k_0 + kw)\phi(w) \text{ and } f_2(w) = \frac{1}{q}\Phi(-k_0 + kw)\phi(w),$$

where $\Phi(.)$ and $\phi(.)$ is the cumulative distribution function ($CDF$) and probability density function ($PDF$) of the standard normal distribution. Under $H_1$, the density function of W can be expressed as mixture combination of two density functions

$$f_W(w) = pf_1(w - w_1) + qf_2(w - w_2).$$

Under $H_1, V_{n_1} \sim N(\lambda, 1)$, therefore, we can calculate $p$ as

$$p = Pr(V_{n_1} > 0) = Pr(V_{n_1} - \lambda > -\lambda) = 1 - \Phi(-\lambda) = \Phi(\lambda).$$

Let $W_1 \sim f_1$ and $W_2 \sim f_2$ be two random variables with PDFs $f_1$ and $f_2$, respectively. The means and variances of $W_1$ and $W_2$ under $H_1$ were shown to be:

$$\mu_1 = \frac{\Lambda}{p}, {\sigma_1}^2 = 1 - \lambda\eta\mu_1 - {\mu_1}^2,$$

$$\mu_2 = \frac{\Lambda}{q}, {\sigma_2}^2 = 1 + \lambda\eta\mu_2 - {\mu_1}^2,$$

respectively, where

$$\Lambda = \frac{\eta}{\sqrt{2\pi}} e(-\frac{1}{2}\lambda^2).$$

Therefore, the mean and variance of W can be derived as

$$\mu_W = 2\Lambda + [pw_1 + qw_2],$$

$$\sigma_W^2 = (1 - 4\Lambda^2) + 2\Lambda[w_1(1-2p) + w_2(1-2q)] + pq(w_1 - w_2)^2,$$

respectively.

At $\lambda = 0$, and under the $H_0$ that $\delta_1 = \delta_2 = 0$ (then $w_1 = w_2 = 0$), the authors showed that $F_W = F_0$, where $F_0$ has density of $f_0(w) = 2\Phi(kw)\phi(w)$ with mean $\mu_0 = \sqrt{\frac{2}{\pi}}\eta$ and variance $\sigma_0^2 = 1 - (\frac{2}{\pi})\eta^2$.

***Type-I error Rate of the Two-Stage Winner Design***

Let $w_\alpha$ be the upper $100\alpha$ percent quantile of $F_0$,

$$w_\alpha = F_W^{-1}(1 - \alpha|H_0) = F_0^{-1}(1 - \alpha),$$

then the one-sided rejection region is

$$\Omega = \{W : W > w_\alpha\},$$

and the Type-I error probability is

$$Pr(W > w_\alpha|H_0) = \alpha.$$

***Normal Approximation***

Even the exact tail probability can be obtained by numerical integration, the authors still proposed a normal approximation approach in the two-stage winner design for the following

reasons: 1) to simplify the calculation of type I error and power using normal approximation rather than using the complex numerical integration; 2) to simplify the calculation on the sample size and to simplify statistical inferences by means of normal approximation; 3) to easily implement the winner design with the normal approximation through their well-understood properties.

Assume $W_1$ and $W_2$ can be approximated by the following normal random variables: $Z_1 \sim N(\mu_1, \sigma_1)$ and $Z_2 \sim N(\mu_2, \sigma_2)$. The density of W can be approximated by

$$f_W(w) \simeq \frac{p}{\sigma_1}\phi(\frac{w - w_1 - \mu_1}{\sigma_1}) + \frac{q}{\sigma_2}\phi(\frac{w - w_2 - \mu_2}{\sigma_2}).$$

***Type I Error rate Using the Normal Approximation***

Under $H_0$, $W^{'}$ $(= \frac{W-\mu_0}{\sigma_0})$. can be approximated by the standard normal distribution. Therefore, the rejection region $\tilde{\Omega}$ can be approximated by a region defined by $\sigma_0 z_\alpha + \mu_0$.

$$\begin{aligned}\tilde{\Omega} &= \{W : W > w_\alpha\} = \{W^{'} : W^{'} > w_\alpha^{'} = \frac{w_\alpha - \mu_0}{\sigma_0}\} \\ &\simeq \{W : W > \sigma_0 z_\alpha + \mu_0\}.\end{aligned}$$

The approximated type I error rate can be calculated by

$$1 - \Phi(\frac{w_\alpha - \mu_0}{\sigma_0})$$

***Power and Sample Size using Normal Approximation***

- Assume $\delta_1 = \delta_2 = \delta$

The targeted overall power can be calculated by

$$1-\beta = Pr(W > w_\alpha | H_1) = 1 - F_W(w_\alpha | \delta_1 = \delta_2 = \delta) = 1 - F_0(w_\alpha - w_0),$$

where $w_0 = \sqrt{\frac{n}{2\sigma_Y^2}}\delta$. Let $w_\beta = -(w_\alpha - w_0)$, using the normal approximation, we have

$$\beta = \Phi(-z_\beta) = F_0(-w_\beta) \simeq \Phi(\frac{-w_\beta - \mu_0}{\sigma_0}),$$

where $z_\beta = \Phi^{-1}(1-\beta)$.

$$1-\alpha = \Phi(z_\alpha) = F_0(w_\alpha) \simeq \Phi(\frac{w_\alpha - \mu_0}{\sigma_0}).$$

Therefore,

$$w_\beta \simeq \sigma_0 z_\beta - \mu_0 \text{ and } w_\alpha \simeq \sigma_0 z_\alpha + \mu_0.$$

Since $w_0 = \sqrt{\frac{n}{2\sigma_Y^2}}\delta = w_\alpha + w_\beta$, for the test to have a power of $1-\beta$ with significance level of $\alpha$, the sample size can be estimated by

$$n \simeq 2\sigma_0^2 (z_\alpha + z_\beta)^2 (\frac{\sigma_y}{\delta})^2.$$

- Assume $\delta_1 \neq \delta_2$ and $\delta_1 < \delta_2$

  The overall power can be calculated by

$$1-\beta = Pr(W > w_\alpha | H_1) = 1 - [pF_1(w_\alpha - w_1) + qF_2(w_\alpha - w_2)] = 1 - [p\beta_1 + q\beta_2]$$

  where $\beta_j = F_j(w_\alpha - w_j)$ for $j = 1, 2$.

  Let $-w_{\beta_j} = w_\alpha - w_j$, for $j = 1, 2$. From the normal approximation, $F_j$ can be approximated by $N(\mu_j, \sigma_j)$, for $j = 1, 2$, and $F_0$ can be approximated by $N(\mu_0, \sigma_0)$, we have

$$w_{\beta_j} \simeq \sigma_j z_{\beta_j} - \mu_j \text{ and } w_\alpha \simeq \sigma_0 z_\alpha + \mu_0.$$

Hence

$$w_j = \sqrt{\frac{n}{2\sigma_Y^2}}\delta_j = w_\alpha + w_{\beta_j} \simeq \mu_0 + \sigma_0 z_\alpha + \sigma_j z_{\beta_j} - \mu_j,$$

for $j = 1, 2$. Therefore for the test with a power of $1 - \beta$ and significance level of $\alpha$, the sample size can be estimated by

$$n \simeq 2(\mu_0 + \sigma_0 z_\alpha + \sigma_j z_{\beta_j} - \mu_j)^2 (\frac{\sigma_Y}{\delta_j})^2, j = 1 \text{ or } 2,$$

with constraints

$$\frac{\delta_1}{\delta_2} = \frac{\mu_0 + \sigma_0 z_\alpha + \sigma_1 z_{\beta_1} - \mu_1}{\mu_0 + \sigma_0 z_\alpha + \sigma_2 z_{\beta_2} - \mu_2},$$

and

$$1 - \beta = 1 - (p\beta_1 + q\beta_2).$$

# Chapter 3

# Non-Inferiority Trial under the Two-Stage Winner Design with Pre-specified Non-Inferiority Margin

At the end of last chapter, we briefly reviewed the methodology of the two-stage winner design of Shun et al. (2008). In this chapter, we propose a methodology to conduct a two-stage winner design for non-inferiority trials with a pre-specified non-inferiority margin when there are two experimental treatments and one active control.

## 3.1 Test Statistic and Settings

Consider a two-stage winner design (Shun et al. 2008) with two experimental treatments and one active control of equal sample size. One interim selection is planned to select a more effective ("winner") treatment at information time $\tau = \frac{n_1}{n}$, where $n_1$ ($\leq n$) is the interim sample size. The "winner" treatment is then continued with the active control for testing the non-inferiority hypothesis in the final analysis. Let $\{X_i^{(j)}|i = 1, ..., n_1\}$ denote the interim continuous measurements assumed to be independent identically distributed with $N(\nu_j^X, \sigma_X^2)$, where $j = 0$ represents the active control arm and $j = 1, 2$ denotes the two experimental treatment arms. Let $\{Y_i^{(j)}|i = 1, ..., n\}, j = 0, 1, 2$ denotes the final continuous measurements that are independent identically distributed with $N(\mu_j^Y, \sigma_Y^2)$. Assume the variances $\sigma_X^2$ and $\sigma_Y^2$ are known, and $X_i^{(j)}$ and $Y_i^{(j)}$ are correlated with a correlation $\rho$, i.e., $corr(X_i^{(j)}, Y_i^{(j)}) = \rho$.

Let the interim sample means be $\bar{X}_{n_1}^{(j)} = (1/n_1)\sum_{i=1}^{n_1} X_i^{(j)}, j = 0, 1, 2$, and the final sample means be $\bar{Y}_n^{(j)} = (1/n)\sum_{i=1}^{n} Y_i^{(j)}, j = 0, 1, 2$. Similar to Shun et al. (2008), we use

the following selection procedure at the interim look:

$$\begin{cases} \text{Keep Treatment 1 ,} & \text{when } \bar{X}_{n_1}^{(1)} > \bar{X}_{n_1}^{(2)}, \\ \text{Keep Treatment 2 ,} & \text{when } \bar{X}_{n_1}^{(2)} > \bar{X}_{n_1}^{(1)}. \end{cases}$$

The goal of the design is to select only one experimental treatment at the interim selection.

Let $\Delta_j = \mu_j^Y - \mu_0^Y$ be the unknown treatment effect of the jth experimental treatment versus the active control at the final analysis, where $j = 1$ or 2. We consider the following non-inferiority hypotheses:

$$H_0 : \Delta_1 \leq -\epsilon \text{ and } \Delta_2 \leq -\epsilon, \text{ versus } H_1 : \Delta_1 > -\epsilon \text{ or } \Delta_2 > -\epsilon, \tag{3.1}$$

where $\epsilon$ ($\geq 0$) is the pre-specified non-inferiority margin which can be interpreted as the largest clinically and statistically acceptable difference. Notice that, when $\epsilon = 0$, the hypotheses in (3.1) becomes superiority hypotheses $H_0 : \Delta_1 \leq 0$ and $\Delta_2 \leq 0$, which is more preferred in general than $H_0 : \Delta_1 = 0$ and $\Delta_2 = 0$ in Shun et al. (2008).

The alternative hypothesis $H_1$ states that either treatment 1 or treatment 2 is non-inferior to the active control within an acceptable non-inferiority margin. Assume that the target treatment effect of the trial design is $\Delta_j = \delta_j$, we define the final test statistic as:

$$W^* = \begin{cases} \frac{\bar{Y}_n^{(1)} - \bar{Y}_n^{(0)} + \epsilon}{\sqrt{\frac{2\sigma_Y^2}{n}}} = Z_n^{(1)} + \sqrt{\frac{n}{2\sigma_Y^2}}\epsilon, & \text{if } V_{n_1} > 0, \\ \frac{\bar{Y}_n^{(2)} - \bar{Y}_n^{(0)} + \epsilon}{\sqrt{\frac{2\sigma_Y^2}{n}}} = Z_n^{(2)} + \sqrt{\frac{n}{2\sigma_Y^2}}\epsilon, & \text{if } V_{n_1} < 0, \end{cases}$$

where $V_{n_1} = \sqrt{\frac{n_1}{2\sigma_X^2}}(\bar{X}_{n_1}^{(1)} - \bar{X}_{n_1}^{(2)})$. Note that $V_{n_1}$ follows a normal distribution with mean $\lambda$ and variance 1, where

$$\lambda = \sqrt{\frac{n_1}{2\sigma_X^2}}(\nu_1^X - \nu_2^X).$$

The null hypothesis $H_0$ is rejected when the value of $W^*$ is large, say $W^* > c^*$. To determine $c^*$, we begin by studying the distribution of $W^*$. Specifically, we consider two scenarios: 1) The surrogate endpoint is different from the final endpoint; 2) same endpoint

is used at both the interim and the final analyses.

## 3.2 Distribution of the Test Statistic

### 3.2.1 Different Endpoints at Interim and Final Analyses

When the endpoint at the interim is different from the endpoint at the final analysis, we follow the argument of Shun et al. (2008) and derive the distribution of $W^*$ as:

$$\begin{aligned}
F_{W^*}(w) =& Pr(W^* \leq w; \lambda, \delta_1, \delta_2, \tau, \rho) \\
=& \int_{-\infty}^{w-\sqrt{\frac{n}{2\sigma_Y^2}}\delta_1-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon} \Phi\left(\frac{\lambda+\eta z}{\sqrt{1-\eta^2}}\right)\phi(z)\,dz \\
&+ \int_{-\infty}^{w-\sqrt{\frac{n}{2\sigma_Y^2}}\delta_2-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon} \Phi\left(\frac{-\lambda+\eta z}{\sqrt{1-\eta^2}}\right)\phi(z)\,dz \\
=& p\left[\frac{1}{p}\int_{-\infty}^{w-w_1-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon} \Phi\left(\frac{\lambda+\eta z}{\sqrt{1-\eta^2}}\right)\phi(z)\,dz\right] \\
&+ q\left[\frac{1}{q}\int_{-\infty}^{w-w_2-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon} \Phi\left(\frac{-\lambda+\eta z}{\sqrt{1-\eta^2}}\right)\phi(z)\,dz\right] \\
=& pF_1(w-w_1-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon) + qF_2(w-w_2-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon), \qquad (3.2)
\end{aligned}$$

where $w_j = \sqrt{\frac{n}{2\sigma_Y^2}}\delta_j, j = 1,2;\ \eta = cov(Z_n^{(1)}, V_{n_1}) = \frac{\sqrt{\tau}}{2}\rho;\ p = Pr(V_{n_1} > 0) = \Phi(\lambda), q = 1-p$, and

$$F_j(w) = \int_{-\infty}^{w} f_j(z)dz, \text{for} j = 1,2,$$

where $f_1(w) = \frac{1}{p}\Phi(\frac{\lambda+\eta w}{\sqrt{1-\eta^2}})\phi(w), f_2(w) = \frac{1}{q}\Phi(\frac{-\lambda+\eta w}{\sqrt{1-\eta^2}})\phi(w)$.

The density function of $W^*$ can be written as,

$$f_{W^*}(w) = pf_1(w-w_1-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon) + qf_2(w-w_2-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon).$$

### 3.2.2 Same Endpoints at Interim and Final Analyses

When the endpoints at the interim and the final analyses are the same, it implies $\rho = 1$ and $\eta = \frac{\sqrt{\tau}}{2}\rho = \frac{\sqrt{\tau}}{2}$. As a result,

$$\lambda = \sqrt{\frac{n_1}{2\sigma_X^2}}(\nu_1^X - \nu_2^X) = \sqrt{\frac{n_1}{2\sigma_X^2}}(\delta_1 - \delta_2) = \sqrt{\tau}(w_1 - w_2) = 2\eta(w_1 - w_2),$$

is a function of $\eta, w_1$ and $w_2$.

By replacing the $\lambda$ with $2\eta(w_1 - w_2)$ in Equation 3.2, we derive the distribution of $W^*$ as follows:

$$\begin{aligned} F_{W^*}(w) &= Pr(W^* \le w; \delta_1, \delta_2, \tau) \\ &= \int_{-\infty}^{w-w_1-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon} \Phi\left(\frac{2\eta(w_1 - w_2) + \eta z}{\sqrt{1-\eta^2}}\right)\phi(z)\,dz \\ &\quad + \int_{-\infty}^{w-w_2-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon} \Phi\left(\frac{-2\eta(w_1 - w_2) + \eta z}{\sqrt{1-\eta^2}}\right)\phi(z)\,dz \\ &= pF_1(w - w_1 - \sqrt{\frac{n}{2\sigma_Y^2}}\epsilon) + qF_2(w - w_2 - \sqrt{\frac{n}{2\sigma_Y^2}}\epsilon). \end{aligned} \tag{3.3}$$

The density function of $W^*$ can be written as,

$$f_{W^*}(w) = pf_1(w - w_1 - \sqrt{\frac{n}{2\sigma_Y^2}}\epsilon) + qf_2(w - w_2 - \sqrt{\frac{n}{2\sigma_Y^2}}\epsilon),$$

where,

$$f_1(w) = \frac{1}{p}\Phi(\frac{2\eta(w_1 - w_2) + \eta w}{\sqrt{1-\eta^2}})\phi(w), f_2(w) = \frac{1}{q}\Phi(\frac{-2\eta(w_1 - w_2) + \eta w}{\sqrt{1-\eta^2}})\phi(w),$$

and $p = \Phi(2\eta(w_1 - w_2)), q = 1 - p$.

Notice that the distribution function of $W^*$ takes the same form no matter different or same endpoints are used at the interim and the final analyses (see Equation 3.2 and Equation 3.3). However, the individual $f_j(z)$'s, $j = 1, 2$ in the density functions are defined differently as summarized in (3.4) and (3.5).

$$f_1(w) = \begin{cases} \frac{1}{p}\Phi(\frac{\lambda+\eta w}{\sqrt{1-\eta^2}})\phi(w), & \text{if different endpoints ,} \\ \frac{1}{p}\Phi(\frac{2\eta(w_1-w_2)+\eta w}{\sqrt{1-\eta^2}})\phi(w), & \text{if same endpoints .} \end{cases} \tag{3.4}$$

and

$$f_2(w) = \begin{cases} \frac{1}{q}\Phi(\frac{-\lambda+\eta w}{\sqrt{1-\eta^2}})\phi(w), & \text{if different endpoints,} \\ \frac{1}{q}\Phi(\frac{-2\eta(w_1-w_2)+\eta w}{\sqrt{1-\eta^2}})\phi(w), & \text{if same endpoints.} \end{cases} \tag{3.5}$$

## 3.3 Type I Error

### 3.3.1 Different Endpoints at Interim and Final Analyses

When different endpoints are used at the interim and the final analyses, given $\epsilon$, we consider the type I error of a test with size $\alpha$ associated with $W^*$ as:

$$\begin{aligned} \alpha &= \sup_{\lambda,\delta_1\leq-\epsilon,\delta_2\leq-\epsilon} Pr(W^* > c^*; \lambda, \delta_1, \delta_2, \eta) \\ &= \sup_{\lambda,w_1\leq-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon,w_2\leq-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon} Pr(W^* > c^*; \lambda, w_1, w_2, \eta), \end{aligned} \tag{3.6}$$

where $w_1 = \sqrt{\frac{n}{2\sigma_Y^2}}\delta_1, w_2 = \sqrt{\frac{n}{2\sigma_Y^2}}\delta_2$, and $c^*$ is the critical value of the test. We define a function $\gamma$ as:

$$\begin{aligned} &\gamma(b; \lambda, w_1, w_2, \eta) = \Pr(W^* > b; \lambda, w_1, w_2, \eta) \\ &= 1 - \int_{-\infty}^{b-w_1-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon} \Phi\left(\frac{\lambda+\eta z}{\sqrt{1-\eta^2}}\right)\phi(z)\,dz - \int_{-\infty}^{b-w_2-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon} \Phi\left(\frac{-\lambda+\eta z}{\sqrt{1-\eta^2}}\right)\phi(z)\,dz. \end{aligned}$$

where $b$ is some cut-off value. Equation 3.6 can be rewritten as:

$$\alpha = \sup_{\lambda,w_1\leq-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon,w_2\leq-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon} \gamma(c^*; \lambda, w_1, w_2, \eta).$$

In order to find the supremum of $\gamma(c^*; \lambda, w_1, w_2, \eta)$, we have proven the following two

lemmas in Appendix B.1 and B.2.

**Lemma 3.3.1.** *Given $\lambda, \eta$ and $b$, $\gamma(b; \lambda, w_1, w_2, \eta)$ is monotonically increasing with $w_1$ and $w_2$.*

**Lemma 3.3.2.** *Given $b, \eta$, $w_1$ and $w_2$, the maximum of $\gamma(b; \lambda, w_1, w_2, \eta)$ occurs at $\lambda = \frac{w_1 - w_2}{2\eta}$, if $\rho > 0$.*

Using these two lammas, we can show the following:

**Theorem 3.1.** *For any $0 \leq \eta < 1$ and $b \geq 0$,*

$$\sup_{\lambda, w_1 \leq -\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon, w_2 \leq -\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon} \gamma(b; \lambda, w_1, w_2, \eta) = \gamma(b; \lambda = 0, w_1 = -\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon, w_2 = -\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon, \eta).$$

*Proof.* For any $-\infty < \lambda < \infty$, $w_1 \leq -\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon$ and $w_2 \leq -\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon$, Lemma 3.3.1 shows that

$$\gamma(b; \lambda, w_1, w_2, \eta) \leq \gamma(b; \lambda, -\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon, -\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon, \eta).$$

Lemma 3.3.2 shows that

$$\gamma(b; \lambda, -\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon, -\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon, \eta) \leq \gamma(b; 0, -\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon, -\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon, \eta).$$

Hence we have

$$\gamma(b; \lambda, w_1, w_2, \eta) \leq \gamma(b; 0, -\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon, -\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon, \eta).$$

□

Theorem 3.1 shows that the type I error in Equation 3.6 can be further simplified to:

$$\alpha = \gamma(c^*; \lambda = 0, w_1 = -\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon, w_2 = -\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon, \eta). \tag{3.7}$$

An interesting observation of (3.7) is that, when $\epsilon = 0$, the critical value for the test statistic of the corresponding superiority hypothesis in (3.1) is the same as that for $H_0 : \Delta_1 = 0$ and $\Delta_2 = 0$ in Shun et al. (2008), so are the evaluations of powers and sample sizes.

In other word, we can replace the null hypothesis in Shun et al. (2008) with the preferred one $H_0 : \Delta_1 \leq 0$ and $\Delta_2 \leq 0$.

Since

$$\begin{aligned}
&\gamma(c^*; \lambda = 0, w_1 = -\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon, w_2 = -\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon, \eta) \\
&= 1 - \int_{-\infty}^{c^* - w_1 - \sqrt{\frac{n}{2\sigma_Y^2}}\epsilon} \Phi\left(\frac{\lambda + \eta z}{\sqrt{1-\eta^2}}\right)\phi(z)\,dz - \int_{-\infty}^{c^* - w_2 - \sqrt{\frac{n}{2\sigma_Y^2}}\epsilon} \Phi\left(\frac{-\lambda + \eta z}{\sqrt{1-\eta^2}}\right)\phi(z)\,dz \\
&= 1 - 2\int_{-\infty}^{c^* + \sqrt{\frac{n}{2\sigma_Y^2}}\epsilon - \sqrt{\frac{n}{2\sigma_Y^2}}\epsilon} \Phi\left(\frac{\eta z}{\sqrt{1-\eta^2}}\right)\phi(z)\,dz \\
&= 1 - 2\int_{-\infty}^{c^*} \Phi\left(\frac{\eta z}{\sqrt{1-\eta^2}}\right)\phi(z)\,dz,
\end{aligned}$$

given type I error $\alpha$, $c^*$ can be determined by solving

$$\alpha = 1 - 2\int_{-\infty}^{c^*} \Phi(\frac{\eta z}{\sqrt{1-\eta^2}})\phi(z)dz. \tag{3.8}$$

Equation 3.8 shows that the critical-value $c^*$ determined by the type I error rate $\alpha$ does not depend on $\epsilon$ nor the sample size $n$. It only depends on the parameter $\eta$, which is a function of the information time $\tau$ and the correlation $\rho$ between the interim and final endpoints. This property is not only useful for testing the non-inferiority hypothesis, but also useful for determining the power and sample size when designing a trial.

Table 3.1 shows the critical value $c^*$ for selective $\tau$ and $\rho$ when $\alpha = 0.025$ when a surrogate endpoint is used. For $\eta > 0$, the critical values are all greater than the usual critical value 1.96. If we use this conventional value of 1.96 in the two-stage winner design, the type I error rate will be erroneously inflated, due to the interim selection. In general, the critical value increases with $\rho$ between the interim and the final endpoints. It also increases with $\tau$. If $\eta$ is the same for different combinations of $\tau$ and $\rho$, the resulted critical value will be the same. Figure 3.1 shows the relationship between $\eta$ and $c^*$ at $\alpha = 0.025$ when different endpoints are used at the interim and the final analyses.

Figure 3.1: The association between $\eta$ and $c^*$ at $\alpha = 0.025$

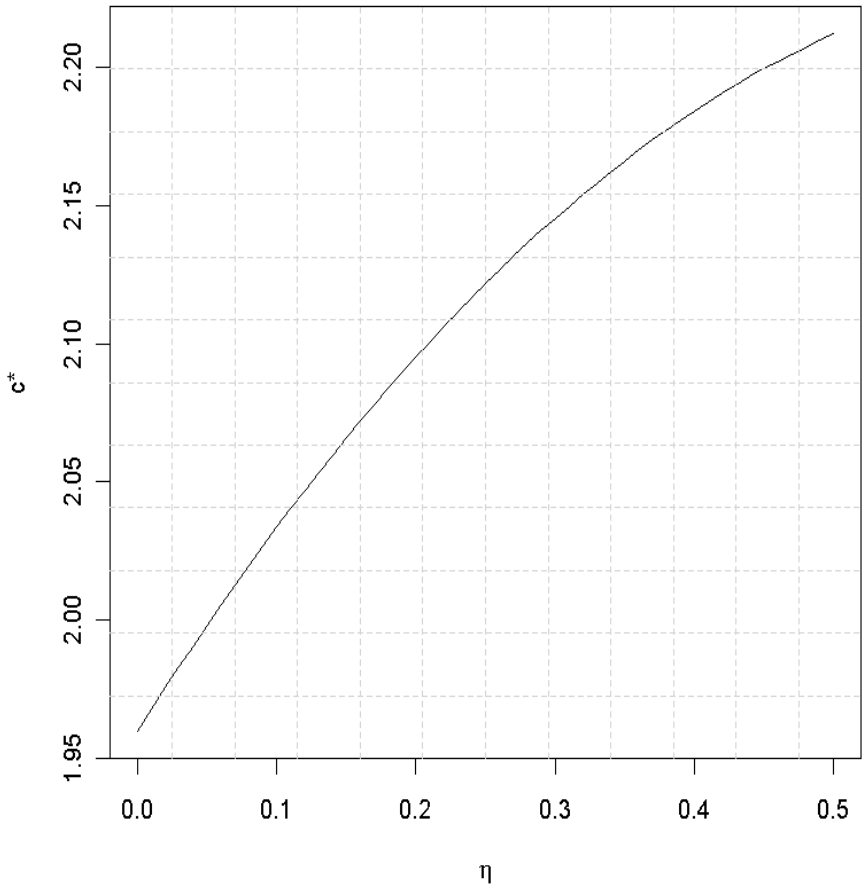


Table 3.1: Critical value of $c^*$ for two-stage winner designs with two experimental and an active control arms by $\eta$ at $\alpha = 0.025$ when different endpoints at interim and final analyses.

| | $(\tau, \rho)$ | | | | |
|---|---|---|---|---|---|
| | $(0.25, 0)$ | $(0.33, 0)$ | $(0.5, 0)$ | $(0.75, 0)$ | $(1, 0)$ |
| $\eta$ | 0 | 0 | 0 | 0 | 0 |
| $c^*$ | 1.960 | 1.960 | 1.960 | 1.960 | 1.960 |
| | $(\tau, \rho)$ | | | | |
| | $(0.25, 0.2)$ | $(0.33, 0.2)$ | $(0.5, 0.2)$ | $(0.75, 0.2)$ | $(1, 0.2)$ |
| $\eta$ | 0.05 | 0.057 | 0.071 | 0.087 | 0.100 |
| $c^*$ | 1.998 | 2.004 | 2.013 | 2.024 | 2.034 |
| | $(\tau, \rho)$ | | | | |
| | $(0.25, 0.5)$ | $(0.33, 0.5)$ | $(0.5, 0.5)$ | $(0.75, 0.5)$ | $(1, 0.5)$ |
| $\eta$ | 0.125 | 0.144 | 0.177 | 0.217 | **0.25** |
| $c^*$ | 2.050 | 2.062 | 2.082 | 2.104 | **2.122** |
| | $(\tau, \rho)$ | | | | |
| | $(0.25, 0.8)$ | $(0.33, 0.8)$ | $(0.5, 0.8)$ | $(0.75, 0.8)$ | $(1, 0.8)$ |
| $\eta$ | 0.2 | 0.300 | 0.283 | 0.346 | 0.400 |
| $c^*$ | 2.095 | 2.111 | 2.138 | 2.165 | 2.184 |

With given $\tau$ and $\rho$, $\eta = \frac{\sqrt{\tau}}{2}\rho$.

### 3.3.2 Same Endpoints at Interim and Final Analyses

When same endpoints are used at the interim and the final analyses, given $\epsilon$, we consider the type I error of a test with size $\alpha$ associated with $W^*$ as:

$$\begin{aligned}\alpha &= \sup_{\delta_1 \le -\epsilon, \delta_2 \le -\epsilon} Pr(W^* > c^*; \delta_1, \delta_2, \eta) \\ &= \sup_{w_1 \le -\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon, w_2 \le -\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon} Pr(W^* > c^*; w_1, w_2, \eta), \end{aligned} \tag{3.9}$$

where $w_1 = \sqrt{\frac{n}{2\sigma_Y^2}}\delta_1, w_2 = \sqrt{\frac{n}{2\sigma_Y^2}}\delta_2$, and $c^*$ is the critical value of the test.

We define a function $\gamma$ as:

$$\begin{aligned}&\gamma(b; w_1, w_2, \eta) = \Pr(W^* > b; w_1, w_2, \eta) \\ &= 1 - \int_{-\infty}^{b - w_1 - \sqrt{\frac{n}{2\sigma_Y^2}}\epsilon} \Phi\left(\frac{2\eta(w_1 - w_2) + \eta z}{\sqrt{1 - \eta^2}}\right) \phi(z)\, dz \\ &\quad - \int_{-\infty}^{b - w_2 - \sqrt{\frac{n}{2\sigma_Y^2}}\epsilon} \Phi\left(\frac{-2\eta(w_1 - w_2) + \eta z}{\sqrt{1 - \eta^2}}\right) \phi(z)\, dz,\end{aligned}$$

where b is some cut-off value. The type I error rate in Equation 3.9 can be rewritten as:

$$\alpha = \sup_{w_1 \leq -\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon, w_2 \leq -\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon} \gamma(c^*; w_1, w_2, \eta).$$

In order to find the supremum of $\gamma(c^*; w_1, w_2, \eta)$, we have proven the following lemma in Appendix B.3

**Lemma 3.3.3.** *For any* $0 \leq \eta < 1$, *and* $0 \leq u_1 \leq u_2$,

$$\begin{aligned}
&\int_{-\infty}^{u_1} \Phi\left(\frac{2\eta(u_2-u_1)+\eta z}{\sqrt{1-\eta^2}}\right)\phi(z)\,dz + \int_{-\infty}^{u_2} \Phi\left(\frac{-2\eta(u_2-u_1)+\eta z}{\sqrt{1-\eta^2}}\right)\phi(z)\,dz \\
&\quad \geq 2\int_{-\infty}^{u_1} \Phi\left(\frac{\eta z}{\sqrt{1-\eta^2}}\right)\phi(z)\,dz
\end{aligned}$$

Using Lemma 3.3.3, we can show the following

**Theorem 3.2.** *For any* $0 \leq \eta < 1$, *and* $b \geq 0$,

$$\sup_{w_1 \leq -\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon, w_2 \leq -\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon} \gamma(b; w_1, w_2, \eta) = \gamma(b; -\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon, -\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon, \eta).$$

*Proof.* For $w_1 \leq -\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon$ and $w_2 \leq -\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon$, without loss of generality, assume that $w_1 \geq w_2$, we have $u_1 = b - w_1 - \sqrt{\frac{n}{2\sigma_Y^2}}\epsilon \geq 0$, $u_2 = b - w_2 - \sqrt{\frac{n}{2\sigma_Y^2}}\epsilon \geq 0$, and $0 \leq u_1 \leq u_2$. By Lemma 3.3.3,

$$\begin{aligned}
\gamma(b; w_1, w_2, \eta) &= 1 - \int_{-\infty}^{b-w_1-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon} \Phi\left(\frac{2\eta(w_1-w_2)+\eta z}{\sqrt{1-\eta^2}}\right)\phi(z)\,dz \\
&\qquad - \int_{-\infty}^{b-w_2-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon} \Phi\left(\frac{-2\eta(w_1-w_2)+\eta z}{\sqrt{1-\eta^2}}\right)\phi(z)\,dz \\
&= 1 - \int_{-\infty}^{u_1} \Phi\left(\frac{2\eta(u_2-u_1)+\eta z}{\sqrt{1-\eta^2}}\right)\phi(z)\,dz - \int_{-\infty}^{u_2} \Phi\left(\frac{-2\eta(u_2-u_1)+\eta z}{\sqrt{1-\eta^2}}\right)\phi(z)\,dz \\
&\leq 1 - 2\int_{-\infty}^{u_1} \Phi\left(\frac{\eta z}{\sqrt{1-\eta^2}}\right)\phi(z)\,dz \\
&\leq 1 - 2\int_{-\infty}^{b} \Phi\left(\frac{\eta z}{\sqrt{1-\eta^2}}\right)\phi(z)\,dz = \gamma(b; -\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon, -\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon, \eta)
\end{aligned}$$

□

Table 3.2: Critical value of $c^*$ for the 2-stage winner designs with two experimental and an active control arms by $\tau$ at $\alpha = 0.025$ when same endpoints at interim and final analyses.

| | $\tau$ | | | | |
|---|---|---|---|---|---|
| | 0.25 | 0.33 | 0.5 | 0.75 | 1 |
| $\eta$ | **0.25** | 0.287 | 0.354 | 0.433 | 0.500 |
| $c^*$ | **2.122** | 2.140 | 2.168 | 2.195 | 2.212 |

With given information time $\tau$, $\eta = \frac{\sqrt{\tau}}{2}$.

Theorem 3.2 shows that the type I error in Equation 3.9 can be further simplified to

$$\alpha = \gamma(c^*; w_1 = -\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon, w_2 = -\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon, \eta) = 1 - 2\int_{-\infty}^{c^*} \Phi(\frac{\eta z}{\sqrt{1-\eta^2}})\phi(z)dz.$$

Hence given type I error rate $\alpha$, the critical value $c^*$ can be determined by solving

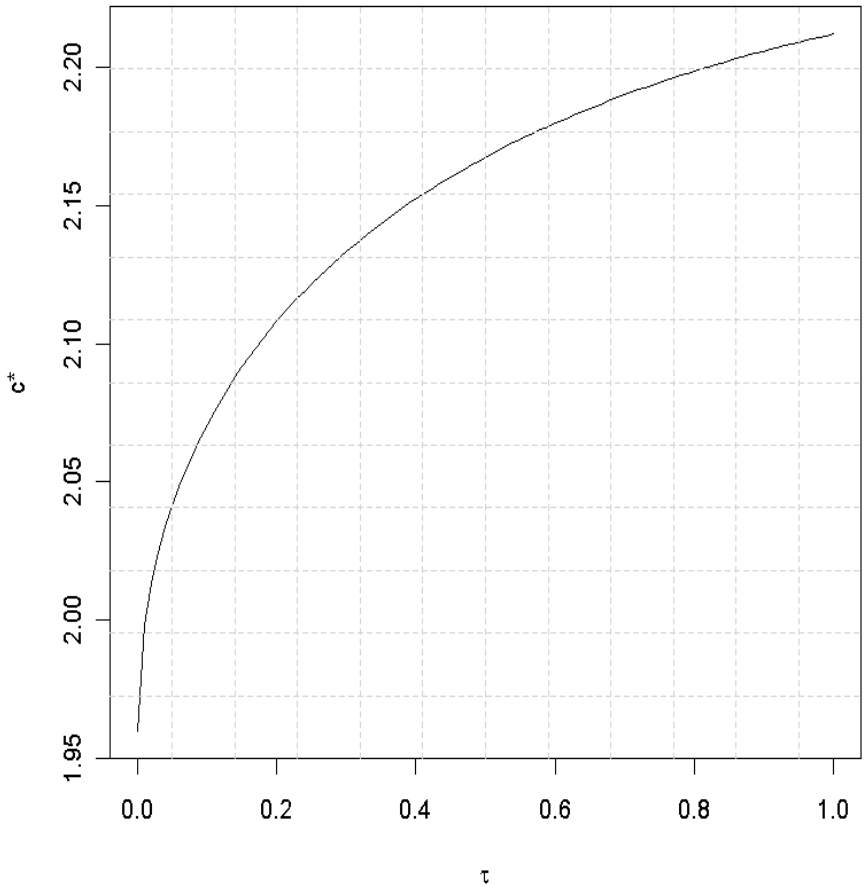
$$\alpha = 1 - 2\int_{-\infty}^{c^*} \Phi(\frac{\eta z}{\sqrt{1-\eta^2}})\phi(z)dz. \tag{3.10}$$

Notice that Equations 3.8 and 3.10 are exactly the same. This implies that the critical value $c^*$ is the same regardless using same or different endpoints at the interim and the final analyses. The critical value $c^*$ does not depend on $\epsilon$ nor $n$. However, $\eta$ only depends on $\tau$ when the same endpoint is used at the interim and final analyses.

Table 3.2 shows the critical value $c^*$ for selected $\tau$ when $\alpha = 0.025$ when same endpoints are used at the interim and the final analyses. For $\tau > 0$, the critical values are all greater than 1.96, a commonly used critical value. If we use this conventional value 1.96 in the two-stage winner design, the type I error rate would be erroneously inflated. The inflation increases as $\tau$ increases.

When a surrogate endpoint is not used, if the given $\tau$ resulted in the same $\eta$ as when a surrogate endpoint is used, the corresponding critical values $c^*$ will be the same. For example, in Table 3.2, when $\tau = 0.25$, the corresponding $\eta = 0.25$, the critical value $c^* = 2.122$ is exactly the same for $\eta = 0.25$ in Table 3.1. Figure 3.2 shows the relationship between the $\tau$ and $c^*$ at $\alpha = 0.025$.

Figure 3.2: Information time $\tau$ against $c^*$ when $\alpha = 0.025$.



## 3.4 Power and Sample Size Calculation

In this section, we discuss how the power and sample size can be calculated. Note that the power we discuss here is the power of rejecting the non-inferiority hypothesis $H_0$, not the probability of selecting the right winner.

### 3.4.1 Power

#### 3.4.1.1 Different Endpoints at Interim and Final Analyses

When different endpoints are used at the interim and the final analyses, we define the power $1-\beta$ as follows:

$$\begin{aligned}
1-\beta &= \Pr(W^* > c^*; H_1) \\
&= \gamma(c^*; \lambda, w_1, w_2, \eta) \\
&= 1 - \int_{-\infty}^{c^* - w_1 - \sqrt{\frac{n}{2\sigma_Y^2}}\epsilon} \Phi\left(\frac{\lambda + \eta z}{\sqrt{1-\eta^2}}\right)\phi(z)\,dz \\
&\quad - \int_{-\infty}^{c^* - w_2 - \sqrt{\frac{n}{2\sigma_Y^2}}\epsilon} \Phi\left(\frac{-\lambda + \eta z}{\sqrt{1-\eta^2}}\right)\phi(z)\,dz. \qquad (3.11)
\end{aligned}$$

In the following paragraphs, we discuss the power function under two scenarios.

- Case I: $\delta_1 = \delta_2 = \delta$.

  When $\delta_1 = \delta_2 = \delta$, the overall targeted power $1-\beta$ can be calculated as

$$\begin{aligned}
1-\beta &= \gamma(c^*; \lambda, w_1 = w_2 = \sqrt{\frac{n}{2\sigma_Y^2}}\delta, \eta) \\
&= 1 - \int_{-\infty}^{c^* - \sqrt{\frac{n}{2\sigma_Y^2}}(\delta+\epsilon)} \Phi\left(\frac{\lambda + \eta z}{\sqrt{1-\eta^2}}\right)\phi(z)\,dz \\
&\quad - \int_{-\infty}^{c^* - \sqrt{\frac{n}{2\sigma_Y^2}}(\delta+\epsilon)} \Phi\left(\frac{-\lambda + \eta z}{\sqrt{1-\eta^2}}\right)\phi(z)\,dz. \qquad (3.12)
\end{aligned}$$

  From Equation 3.12, power is a function of $n$, $c^*, \epsilon$, $\eta, \sigma_Y, \delta$ and $\lambda$. For different combinations of $\delta$ and $\epsilon$, as long as the sum of these two parameters are the same, so is the power. Given $\alpha$ and $\eta$, critical value $c^*$ can be obtained from Equation 3.8.

Table 3.3: Probability of selecting treatment 1, by $\tau$ and $\nu_1^X - \nu_2^X$ when different endpoints at interim and final analyses.

| | $\nu_1^X - \nu_2^X = 0$ | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\tau = 0.25$ | | | | | $\tau = 0.5$ | | | | | $\tau = 0.75$ | | | | |
| n | 100 | 250 | 500 | 1000 | 2000 | 100 | 250 | 500 | 1000 | 2000 | 100 | 250 | 500 | 1000 | 2000 |
| $p$ | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 |
| | $\nu_1^X - \nu_2^X = 0.05$ | | | | | | | | | | | | | | |
| | $\tau = 0.25$ | | | | | $\tau = 0.5$ | | | | | $\tau = 0.75$ | | | | |
| n | 100 | 250 | 500 | 1000 | 2000 | 100 | 250 | 500 | 1000 | 2000 | 100 | 250 | 500 | 1000 | 2000 |
| $p$ | 0.57 | 0.61 | 0.65 | 0.71 | 0.79 | 0.60 | 0.65 | 0.71 | 0.79 | 0.87 | 0.62 | 0.69 | 0.75 | 0.83 | 0.91 |
| | $\nu_1^X - \nu_2^X = 0.1$ | | | | | | | | | | | | | | |
| | $\tau = 0.25$ | | | | | $\tau = 0.5$ | | | | | $\tau = 0.75$ | | | | |
| n | 100 | 250 | 500 | 1000 | 2000 | 100 | 250 | 500 | 1000 | 2000 | 100 | 250 | 500 | 1000 | 2000 |
| $p$ | 0.64 | 0.71 | 0.79 | 0.87 | 0.94 | 0.69 | 0.79 | 0.87 | 0.94 | 0.99 | 0.73 | 0.83 | 0.91 | 0.97 | 1.00 |
| | $\nu_1^X - \nu_2^X = 0.3$ | | | | | | | | | | | | | | |
| | $\tau = 0.25$ | | | | | $\tau = 0.5$ | | | | | $\tau = 0.75$ | | | | |
| n | 100 | 250 | 500 | 1000 | 2000 | 100 | 250 | 500 | 1000 | 2000 | 100 | 250 | 500 | 1000 | 2000 |
| $p$ | 0.86 | 0.95 | 0.99 | 1.00 | 1.00 | 0.93 | 0.99 | 1.00 | 1.00 | 1.00 | 0.97 | 1.00 | 1.00 | 1.00 | 1.00 |

To study the effect of $n$, $\epsilon$ and $\nu_1^X - \nu_2^X$ on power, we show the power curves for $\delta_1 = \delta_2 = \delta$ when $\alpha = 0.025, \tau = 0.25, \rho = 0.8, \sigma_X = \sigma_Y = 1$ in Figure 3.3. When $\nu_{12}$ changes, there is no obviously changes on the power in the figures. This is clear since the final targeted treatment effects are the same ($\delta_1 = \delta_2$), the experimental treatment effects at interim look have minimum impact on the power. As $\epsilon$ increases, power curves shift to the left, suggesting the increase on the power.

To study the effect of $n$, $\tau$ and $\rho$ on the power, we show the power curves for for $\delta_1 = \delta_2 = \delta$ when $\alpha = 0.025, \nu_1^X - \nu_2^X = 0.1, \sigma_X = \sigma_Y = 1, \epsilon = 0.1$ in Figure 3.4. When $\tau$ or $\rho$ increases, there is no visual increase on the power. This is clear because the two experimental treatments have the same effect at the final analysis, the timing of the interim analysis, or the correlation between the interim and the final endpoints have minimum effect on the power.

- Case II: $\delta_1 \neq \delta_2$.

  When $\delta_1 \neq \delta_2$, the power $1 - \beta$ can be expressed as

$$1 - \beta = \gamma(c^*; \lambda, w_1 = \sqrt{\frac{n}{2\sigma_Y^2}}\delta_1, w_2 = \sqrt{\frac{n}{2\sigma_Y^2}}\delta_2, \eta)$$

Figure 3.3: power curves for $\delta_1 = \delta_2 = \delta$ when $\alpha = 0.025, \tau = 0.25, \rho = 0.8$ and $\sigma_X = \sigma_Y = 1$ by $n$, $\epsilon$ and $\nu_{12} = \nu_1^X - \nu_2^X$ when different endpoints at interim and final analyses.

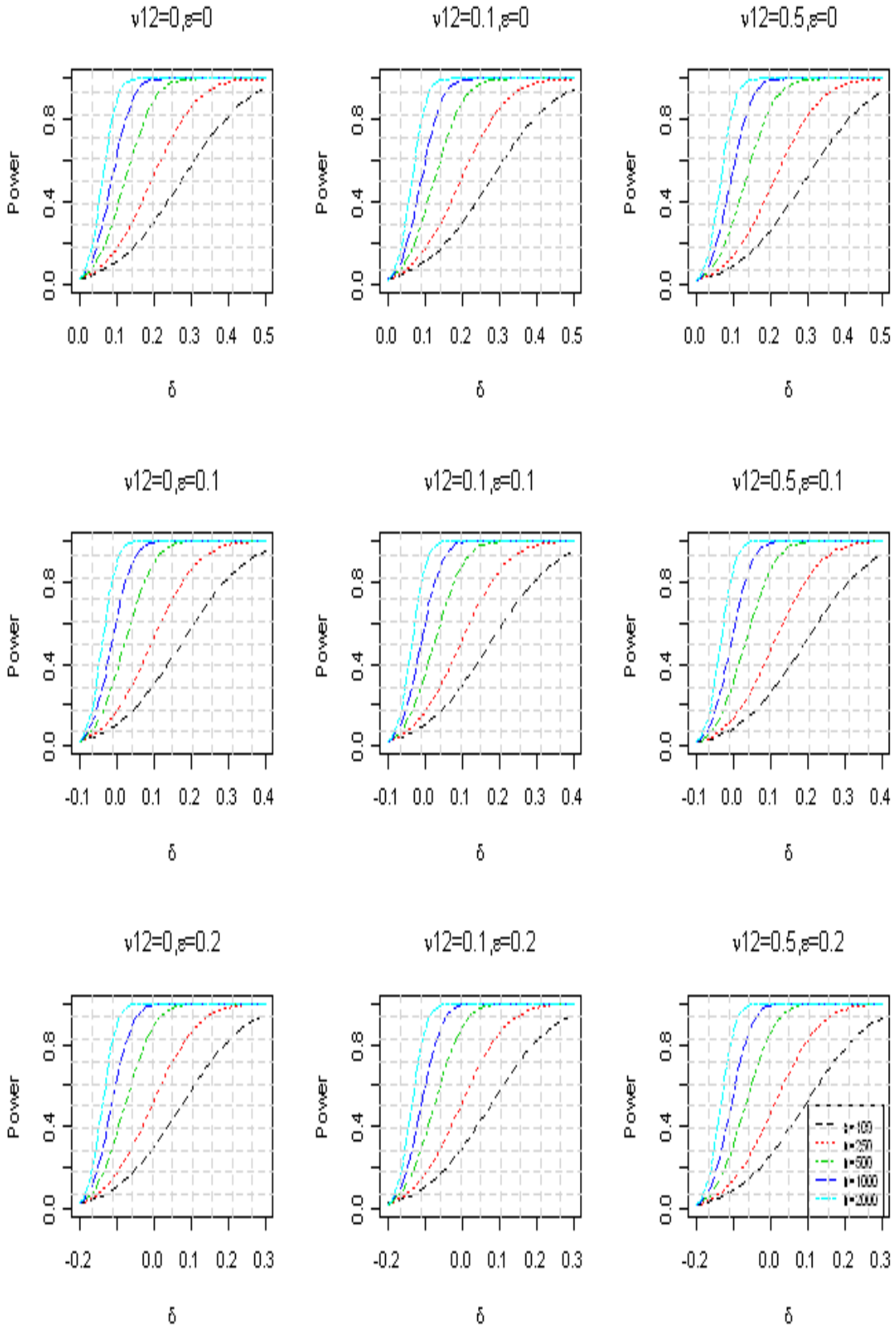
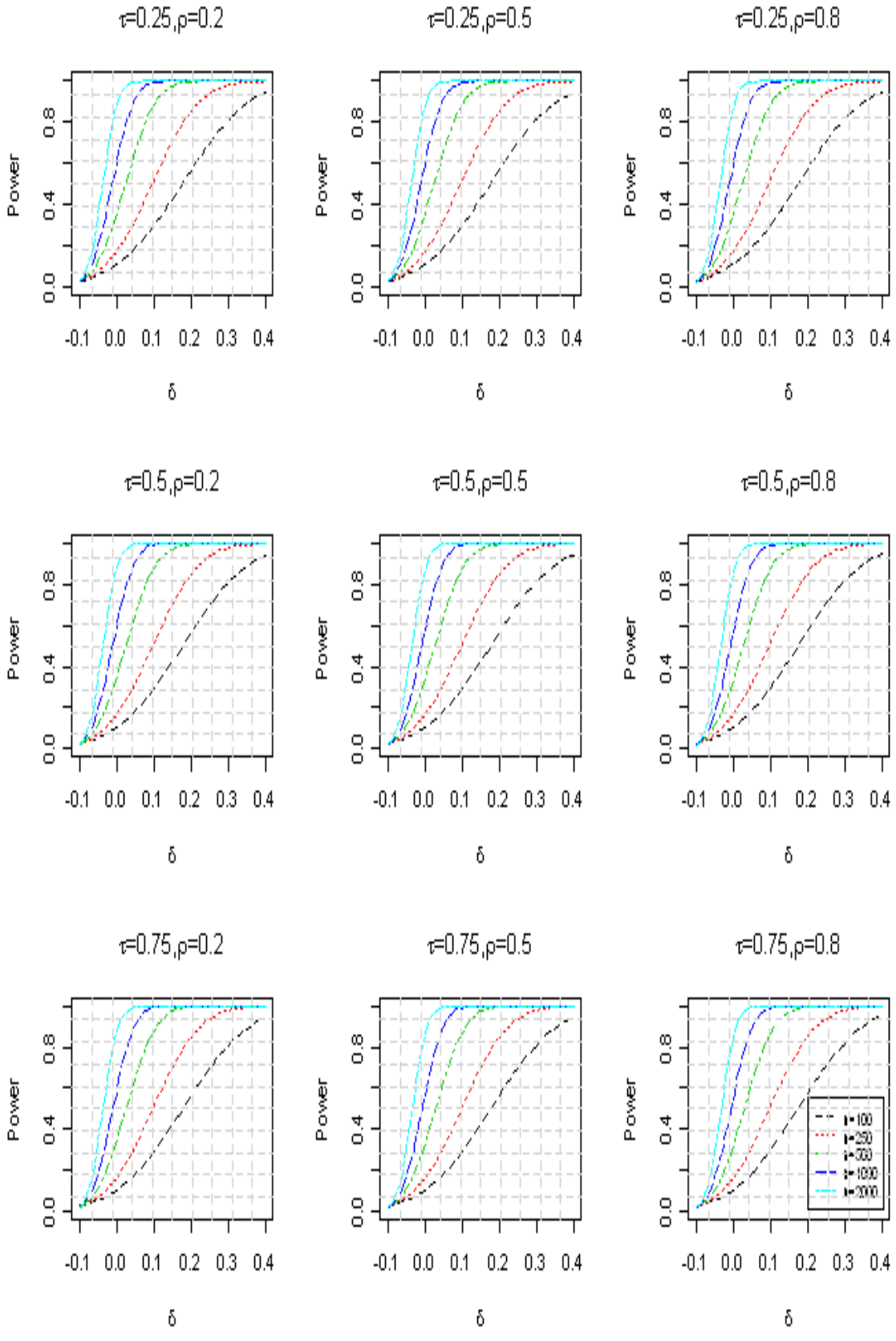


Figure 3.4: Power curves for $\delta_1 = \delta_2 = \delta$ when $\alpha = 0.025, \nu_1^X - \nu_2^X = 0.1, \sigma_X = \sigma_Y = 1, \epsilon = 0.1$ by $n$, $\tau$ and $\rho$ when different endpoints at interim and final analyses.



$$= 1 - \int_{-\infty}^{c^* - \sqrt{\frac{n}{2\sigma_Y^2}}(\delta_1+\epsilon)} \Phi\left(\frac{\lambda+\eta z}{\sqrt{1-\eta^2}}\right)\phi(z)\,dz$$
$$- \int_{-\infty}^{c^* - \sqrt{\frac{n}{2\sigma_Y^2}}(\delta_2+\epsilon)} \Phi\left(\frac{-\lambda+\eta z}{\sqrt{1-\eta^2}}\right)\phi(z)\,dz. \quad (3.13)$$

From Equation 3.13, the power is a function of parameters $c^*, \delta_1, \delta_2, \lambda, \sigma_Y, \eta, \epsilon$ and $n$. Given $\alpha$ and $\eta$, critical value $c^*$ can be obtained from Equation 3.8.

To study the effect of $n$ and $\nu_{12} = \nu_1^X - \nu_2^X$ on the power, we show the contour of power surface for $\delta_1 \neq \delta_2$ when $\tau = 0.25, \rho = 0.8, \epsilon = 0.1, \sigma_X = \sigma_Y = 1$ and $\alpha = 0.025$ in Figure 3.5. When $\nu_1^X - \nu_2^X$ increases, power increases. Table 3.3 provide the probability of selecting treatment 1, by $\tau$ and $\nu_1^X - \nu_2^X$. When $n = 100$, as the $\nu_1^X - \nu_2^X$ increase to 0.3, the probability of selecting treatment 1 at the interim selection is at least 0.86. Therefore, the bottom 3 figures in Figure 3.5 show that $\delta_2$ has a minimal impact on power. As $n$ increases, power surfaces shift to the left suggesting the increase on power.

To study the effect of $\tau$ and $\rho$ on the power, we show the contour of power surface for $\delta_1 \neq \delta_2$ when $n = 250, \nu_1^X - \nu_2^X = 0.1, \epsilon = 0.1, \sigma_X = \sigma_Y = 1$ and $\alpha = 0.025$ in Figure 3.6. As seen in Table 3.3, when $n = 250, \nu_1^X - \nu_2^X = 0.1$, the probability of selecting treatment 1 increases from 0.71 to 0.83, as $\tau$ increases from 0.25 to 0.75. As a result, the power increases when $\tau$ increases. On the other hand, when $\rho$ increases, it have minimum impact on the power.

Figure 3.6 and Figure 3.7 are similar except for the selected values of $\nu_1^X - \nu_2^X$ ($\nu_1^X - \nu_2^X = 0.1$ in Figure 3.6, and 0.3 in Figure 3.7. Comparing these two figures, we can see that the power surface shifts to the left when $\nu_1^X - \nu_2^X = 0.3$, which implies the increase on the power. Furthermore, when $\nu_1^X - \nu_2^X = 0.3$, according to Table 3.3, we have at least 90% of probability to select treatment 1. Therefore, the power depends much more on $\delta_1$ then $\delta_2$ in Figure 3.7.

To study the effect of $\epsilon$ and $\eta$ on power, we show the contour of power surface for $\delta_1 \neq \delta_2$ when $\nu_1^X - \nu_2^X = 0.1, n = 250, \sigma_X = \sigma_Y = 1$ and $\alpha = 0.025$ in Figure 3.8. When $\eta$ increases, the surface changes only a little bit. Since $\eta$ is a function of $\tau$ and

Figure 3.5: Contour of power surface for $\delta_1 \neq \delta_2$ when $\tau = 0.25, \rho = 0.8, \epsilon = 0.1, \sigma_X = \sigma_Y = 1$ and $\alpha = 0.025$ by $n$, and $\nu_{12} = \nu_1^X - \nu_2^X$ when different endpoints at interim and final analyses.

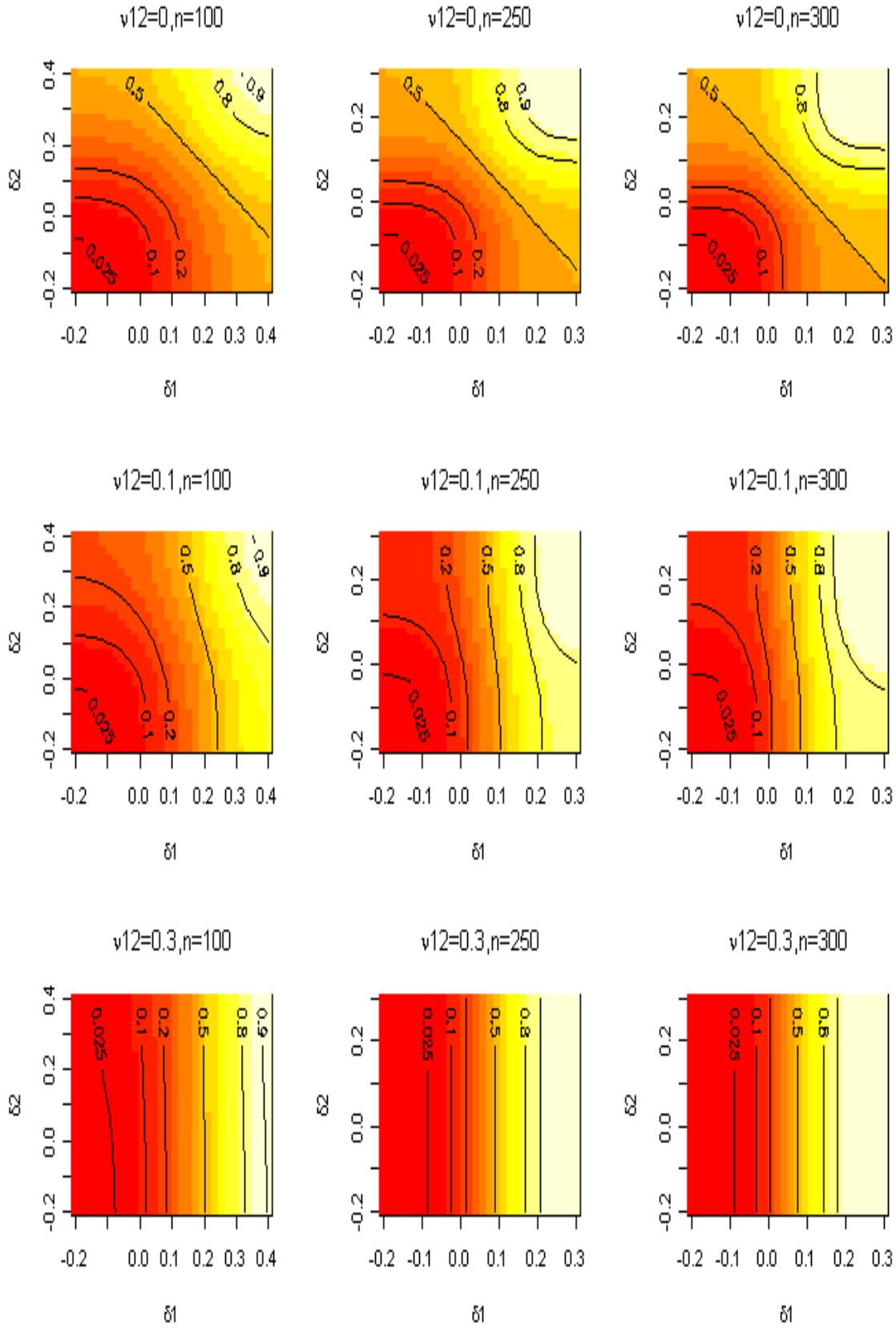


Figure 3.6: Contour of power surface for $\delta_1 \neq \delta_2$ when $\nu_1^X - \nu_2^X = 0.1, n = 250, \epsilon = 0.1, \sigma_X = \sigma_Y = 1$ and $\alpha = 0.025$ by $\tau$ and $\rho$ when different endpoints at interim and final analyses.

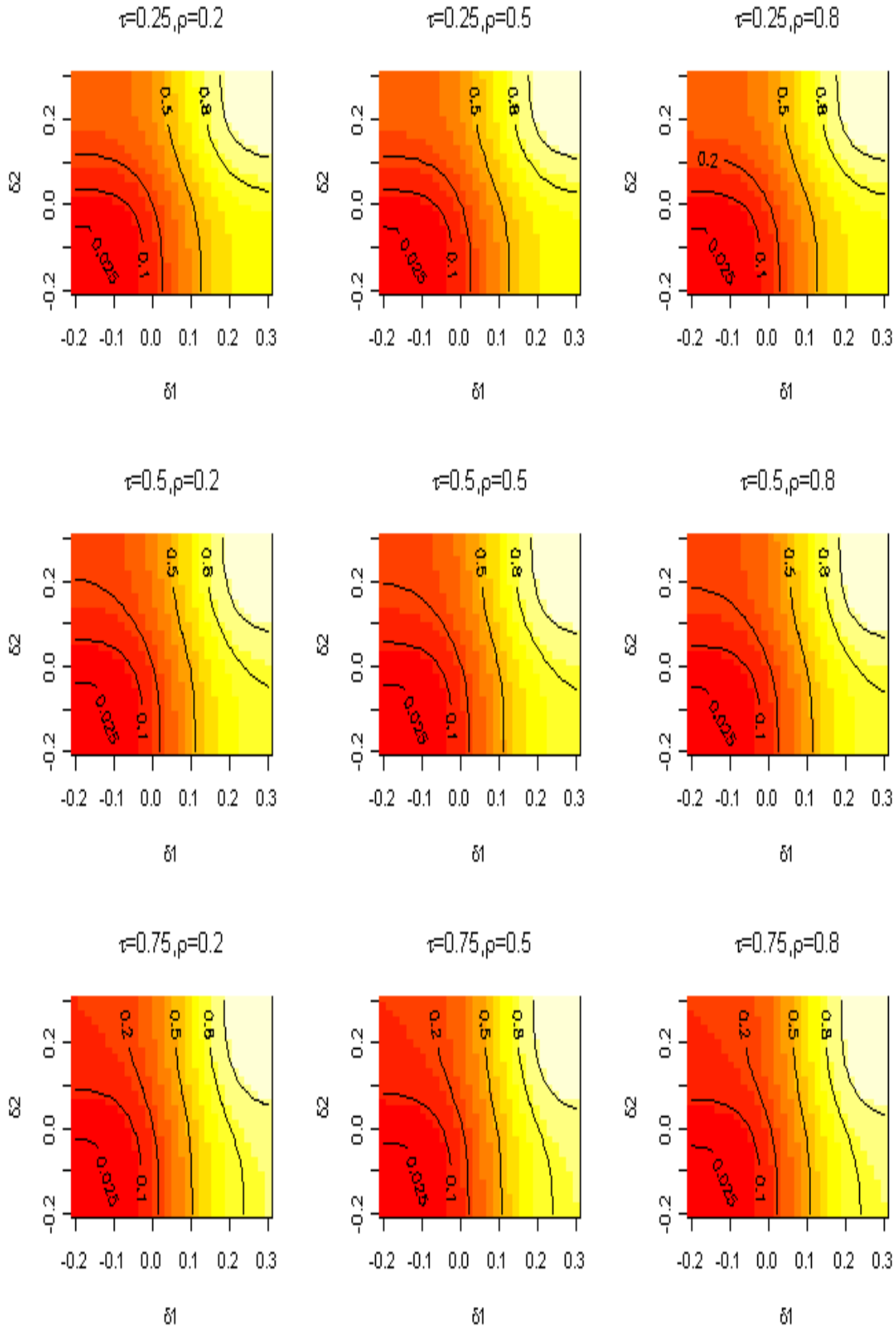
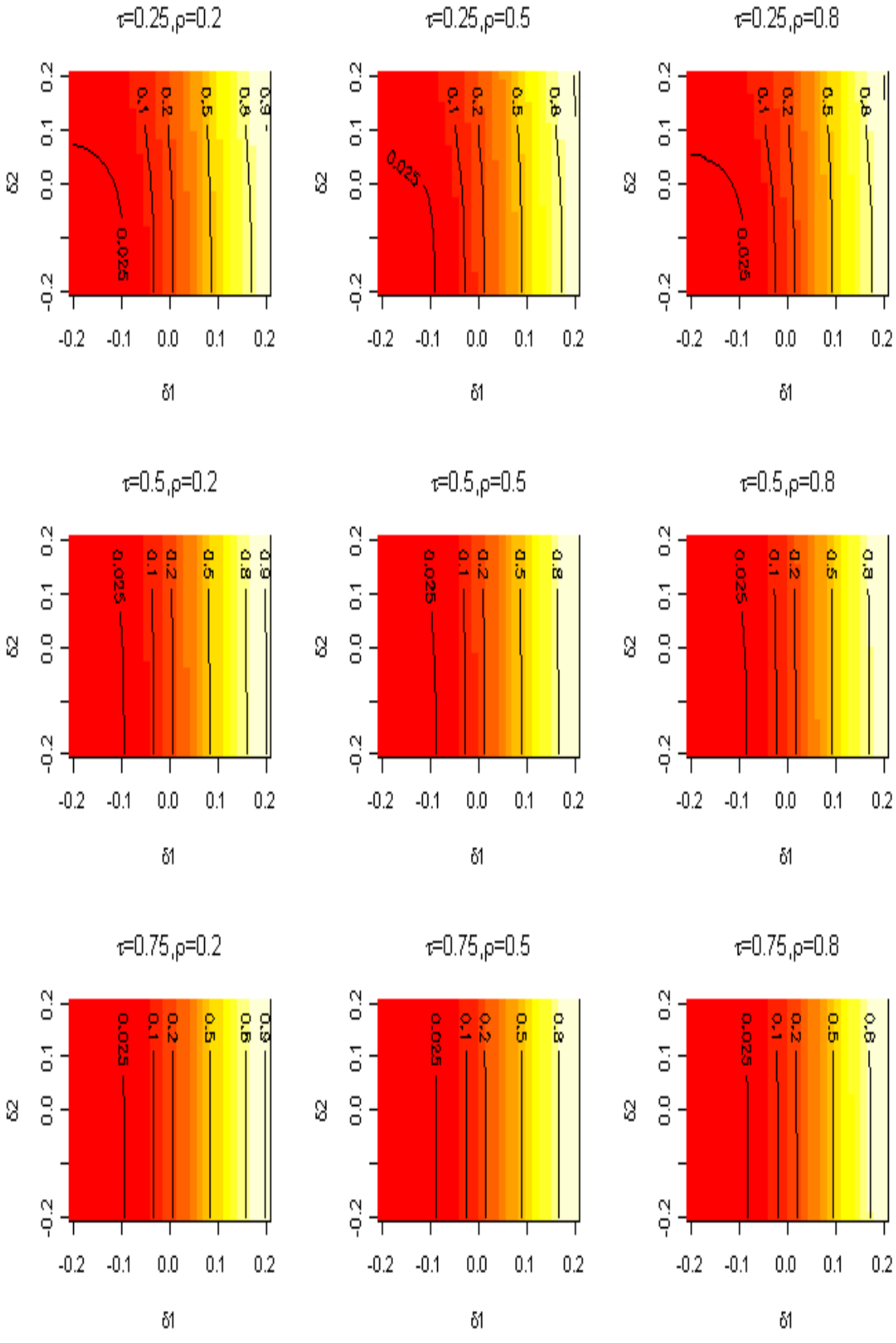


Figure 3.7: Contour of power surface for $\delta_1 \neq \delta_2$ when $\nu_1^X - \nu_2^X = 0.3, n = 250, \epsilon = 0.1, \sigma_X = \sigma_Y = 1$ and $\alpha = 0.025$ by $\tau$ and $\rho$ when different endpoints at interim and final analyses.



$\rho$, it hard for us to see its individual effect on the power. On the other hand, power increases as the $\epsilon$ increases.

#### 3.4.1.2 Same Endpoints at Interim and Final Analyses

When the interim and final endpoints are the same, we define the power $1-\beta$ as follows:

$$\begin{aligned}
1-\beta &= \Pr(W^* > c^*; H_1) \\
&= \gamma(c^*; w_1, w_2, \eta) \\
&= 1 - \int_{-\infty}^{c^*-w_1-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon} \Phi\left(\frac{2\eta(w_1-w_2)+\eta z}{\sqrt{1-\eta^2}}\right)\phi(z)\,dz \\
&\quad - \int_{-\infty}^{c^*-w_2-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon} \Phi\left(\frac{-2\eta(w_1-w_2)+\eta z}{\sqrt{1-\eta^2}}\right)\phi(z)\,dz. \qquad (3.14)
\end{aligned}$$

From Equation 3.14, we know the power is a function of $c^*, \delta$'s, $\sigma_Y, \epsilon$, $n$ and $\eta$ or $\tau$. Given $\alpha$ and $\eta$, critical value $c^*$ can be obtained from Equation 3.10.

- Case I: $\delta_1 = \delta_2 = \delta$.

  When $\delta_1 = \delta_2 = \delta$, the overall power $1-\beta$ can be calculated as

$$1-\beta = 1 - 2\int_{-\infty}^{c^*-\sqrt{\frac{n}{2\sigma_Y^2}}(\delta+\epsilon)} \Phi\left(\frac{\eta z}{\sqrt{1-\eta^2}}\right)\phi(z)\,dz. \qquad (3.15)$$

  As seen in Equation 3.15, for different combinations of $\delta$ and $\epsilon$, as long as the sum of these two parameters are the same, so is the power.

  To study the effect of $\tau, n$ and $\epsilon$ on the power, we show the power curves for $\delta_1 = \delta_2 = \delta$ when $\alpha = 0.025$ and $\sigma_Y = 1$ in Figure 3.9. Obviously, power increases with $n$ and $\epsilon$. For same $\epsilon$, power remains the same for all choices of $\tau$. This is because when $\delta_1 = \delta_2$ and the endpoint at the interim and the final analysis are the same, it does not matter when and which treatment is selected at the interim.

- Case II: $\delta_1 \neq \delta_2$.

Figure 3.8: Contour of power surface $\delta_1 \neq \delta_2$ when $\nu_1^X - \nu_2^X = 0.1, n = 250, \sigma_X = \sigma_Y = 1$ and $\alpha = 0.025$ by $\epsilon$ and $\eta$ when different endpoints at interim and final analyses.

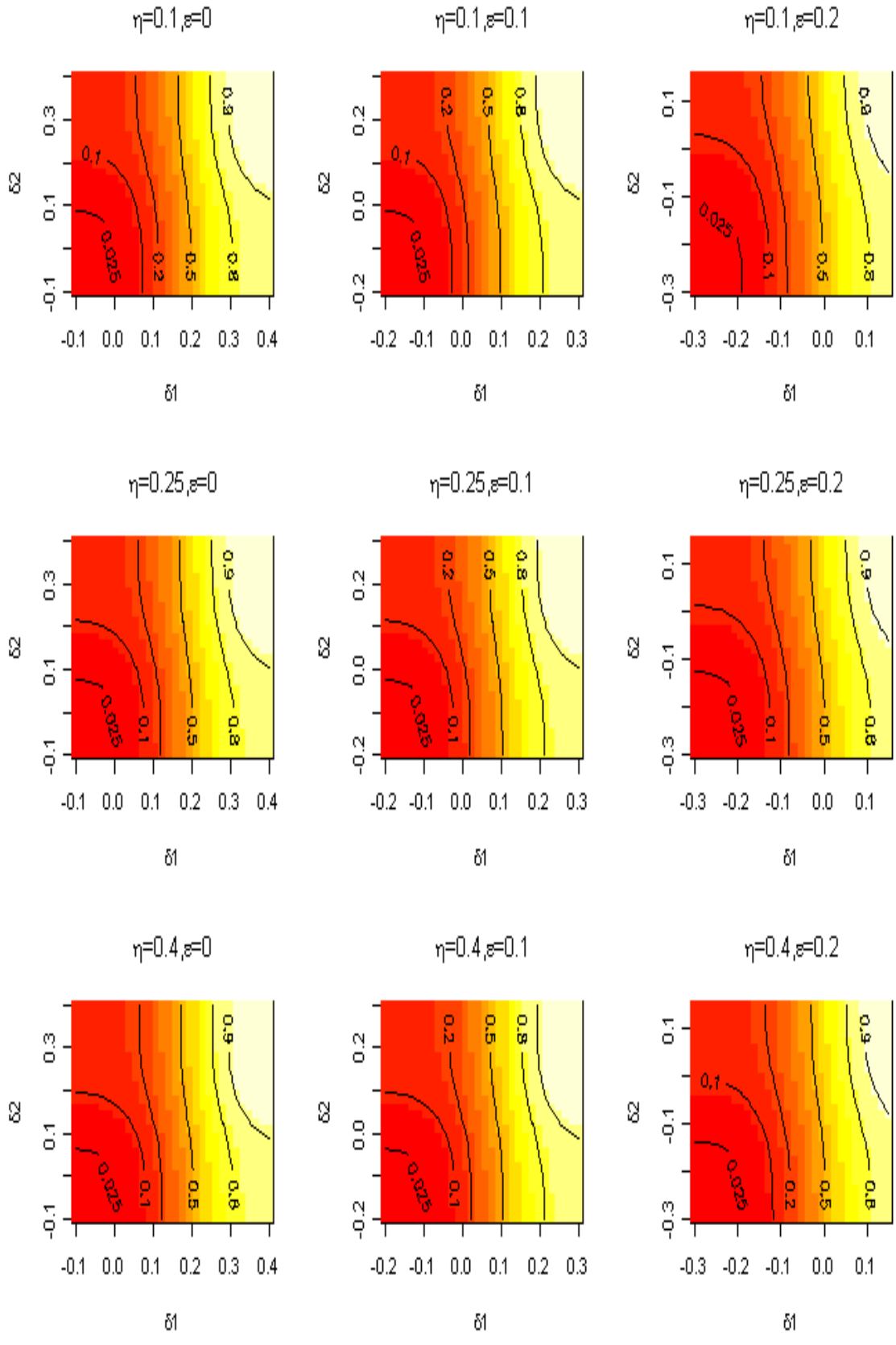
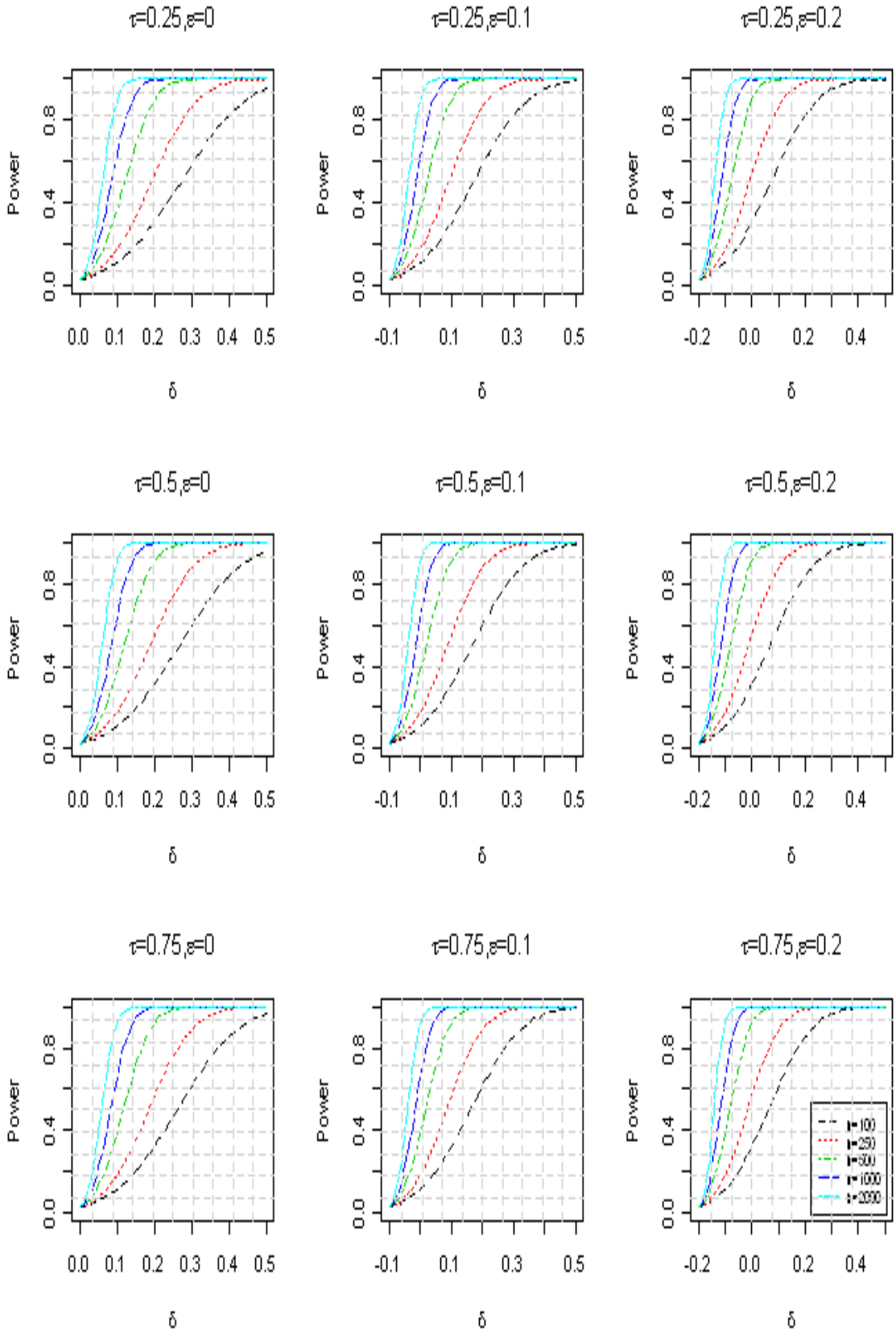


Figure 3.9: Power curve for $\delta_1 = \delta_2 = \delta$ when $\alpha = 0.025$ and $\sigma_Y = 1$ by $\tau, n$ and $\epsilon$ when same endpoints at interim and final analyses.



When $\delta_1 \neq \delta_2$ , the power $1-\beta$ can be expressed as

$$1-\beta = 1 - \int_{-\infty}^{c^* - \sqrt{\frac{n}{2\sigma_Y^2}}(\delta_1+\epsilon)} \Phi\left(\frac{2\eta(w_1-w_2)+\eta z}{\sqrt{1-\eta^2}}\right)\phi(z)\,dz - \int_{-\infty}^{c^* - \sqrt{\frac{n}{2\sigma_Y^2}}(\delta_2+\epsilon)} \Phi\left(\frac{-2\eta(w_1-w_2)+\eta z}{\sqrt{1-\eta^2}}\right)\phi(z)\,dz. \tag{3.16}$$

Equation 3.16 shows the power function is symmetric in $\delta_1$ and $\delta_2$. Regardless $\delta_1 > \delta_2$ or $\delta_1 < \delta_2$, as long as the value of the $\delta$'s are the same, so is the power.

To study the effect of $\tau$ and $\epsilon$ on the power, we show the power curves for $\delta_1 \neq \delta_2$ when $n = 250, \sigma_Y = 1$ and $\alpha = 0.025$ in Figure 3.10. As we can see in the figures, power increases as either $\epsilon$ or $\tau$ increases.

To study the effect of $n$ and $\epsilon$ on the power, we show the power curves for $\delta_1 \neq \delta_2$ when $\tau = 0.5, \sigma_Y = 1$ and $\alpha = 0.025$ in Figure 3.11. It is obvious that power increases as either $\epsilon$ or $n$ increases.

### 3.4.2 Sample Size Calculation

#### 3.4.2.1 Different Endpoints at Interim and Final Analyses

- Case I: $\delta_1 = \delta_2 = \delta$.

  When $\delta_1 = \delta_2 = \delta$, we can determine the sample size $n$ per group by solving Equation 3.12, given the $\alpha, \eta, c^*, \lambda, \sigma_Y$, power $1-\beta$ and $\delta$.

  Table 3.4 provides an example of the estimated sample size for $\delta_1 = \delta_2 = \delta$ when $\nu_1^X - \nu_2^X = 0.1, \rho = 0.8, \sigma_X = \sigma_Y = 1, \alpha = 0.025$ with 80% or 90% power by $\tau, \delta$ and $\epsilon$. At given power and $\tau$, sample size decreases as $\epsilon$ or $\delta$ increases. At given power, $\epsilon$ and $\delta$, sample size increases with $\tau$. When the sum of $\delta$ and $\epsilon$ is the same by fixing other parameters, power remains the same. For example, at $\tau = 0.25$ with 80% power, the sample size needed is the same when $\epsilon = 0, \delta = 0.2 (\epsilon + \delta = 0.2)$ and when $\epsilon = 0.1, \delta = 0.1 (\epsilon + \delta = 0.2)$.

- Case II: $\delta_1 \neq \delta_2$.

Figure 3.10: Contour of power surface for $\delta_1 \neq \delta_2$ when $n = 250, \sigma_Y = 1$ and $\alpha = 0.025$ by $\tau$ and $\epsilon$ when same endpoints at interim and final analyses.

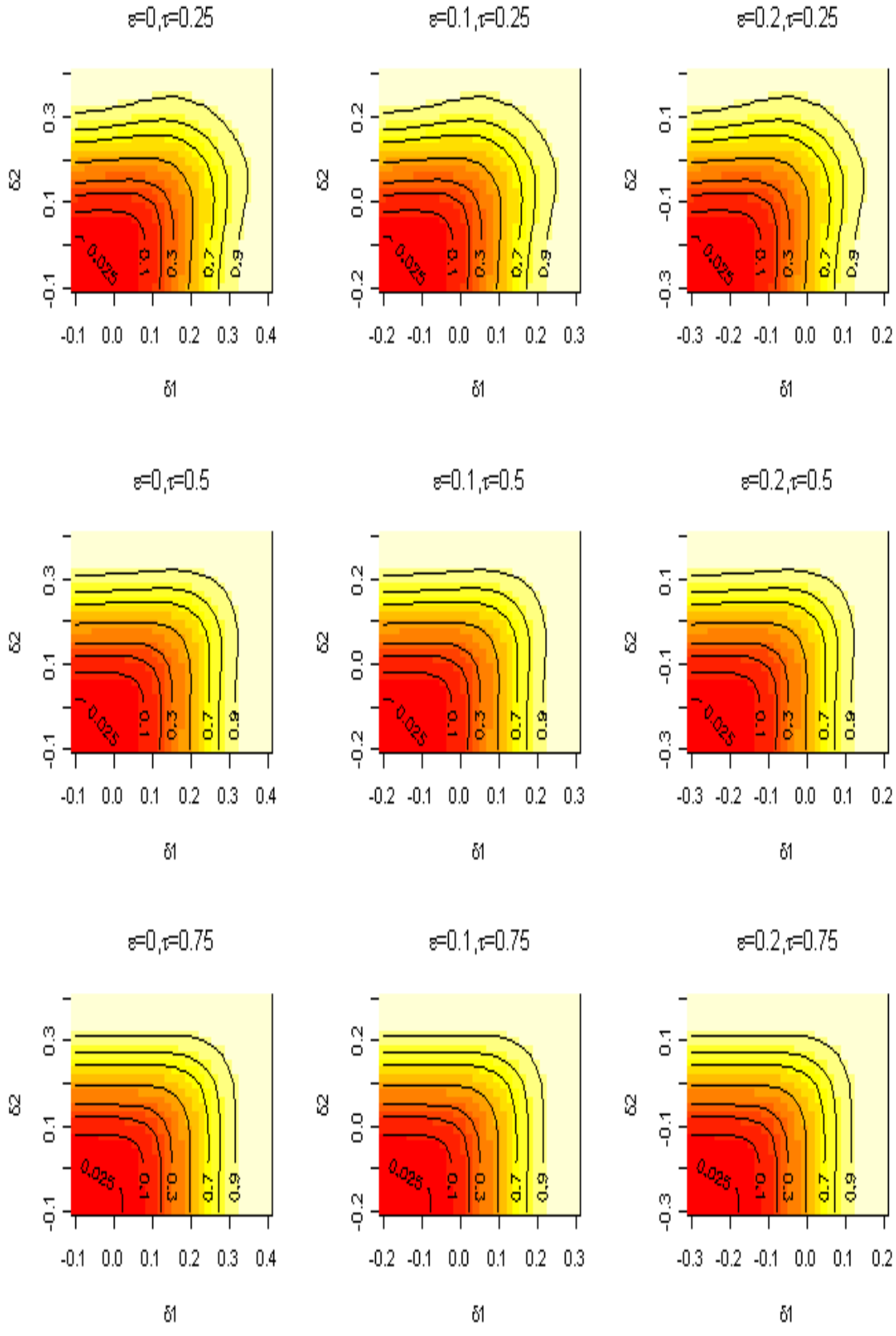


Figure 3.11: Contour of power surface for $\delta_1 \neq \delta_2$ when $\tau = 0.5, \sigma_Y = 1$ and $\alpha = 0.025$ by $n$ and $\epsilon$ when same endpoints at interim and final analyses.

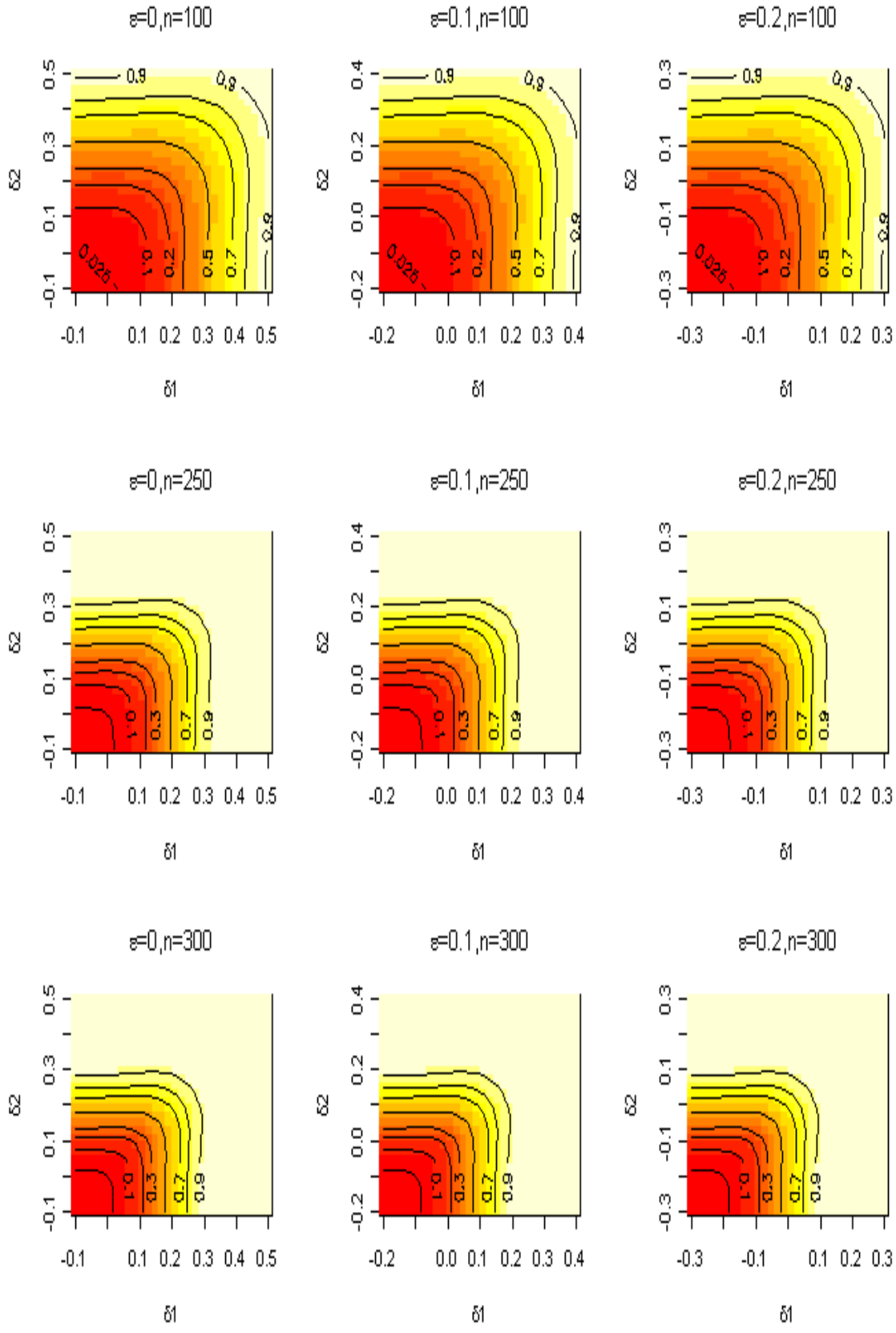


Table 3.4: Estimated sample size for $\delta_1 = \delta_2 = \delta$ when $\nu_1^X - \nu_2^X = 0.1, \rho = 0.8, \sigma_X = \sigma_Y = 1, \alpha = 0.025$ with 80% or 90% power by $\tau, \delta$ and $\epsilon$ when different endpoints at interim and final analyses.

| | $\tau = 0.25, \rho = 0.8 (i.e.\ \eta = 0.2)$ | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\epsilon = 0$ | | | | | $\epsilon = 0.1$ | | | | | $\epsilon = 0.2$ | | | | |
| $\delta$ | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 |
| $1-\beta = 0.8$ | 1658 | 394 | 173 | 97 | 62 | 394 | 173 | 97 | 62 | 43 | 173 | 97 | 62 | 43 | 32 |
| $1-\beta = 0.9$ | 2226 | 530 | 232 | 130 | 83 | 530 | 232 | 130 | 83 | 58 | 232 | 130 | 83 | 58 | 42 |
| | $\tau = 0.5, \rho = 0.8 (i.e.\ \eta = 0.28)$ | | | | | | | | | | | | | | |
| | $\epsilon = 0$ | | | | | $\epsilon = 0.1$ | | | | | $\epsilon = 0.2$ | | | | |
| $\delta$ | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 |
| $1-\beta = 0.8$ | 1746 | 402 | 173 | 96 | 61 | 402 | 173 | 96 | 61 | 42 | 173 | 96 | 61 | 42 | 31 |
| $1-\beta = 0.9$ | 2321 | 544 | 233 | 129 | 82 | 544 | 233 | 129 | 82 | 57 | 233 | 129 | 82 | 57 | 42 |
| | $\tau = 0.75, \rho = 0.8 (i.e.\ \eta = 0.35)$ | | | | | | | | | | | | | | |
| | $\epsilon = 0$ | | | | | $\epsilon = 0.1$ | | | | | $\epsilon = 0.2$ | | | | |
| $\delta$ | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 |
| $1-\beta = 0.8$ | 1797 | 413 | 174 | 95 | 60 | 413 | 174 | 95 | 60 | 42 | 174 | 95 | 60 | 42 | 31 |
| $1-\beta = 0.9$ | 2371 | 559 | 235 | 128 | 81 | 559 | 235 | 128 | 81 | 56 | 235 | 128 | 81 | 56 | 41 |

When $\delta_1 \neq \delta_2$, we can use Equation 3.13 to determine the sample size given the $\alpha, \eta, c^*, \lambda, \sigma_Y$, power $1-\beta$ and $\delta$'s.

Table 3.5 provides an example of the estimated sample size for $\delta_1 \neq \delta_2$ when $\nu_1^X - \nu_2^X = 0.1, \rho = 0.8, \sigma_X = \sigma_Y = 1, \alpha = 0.025$ with 80% or 90% power by $\tau, \delta_1, \delta_2$ and $\epsilon$. When the effect of two treatments are close (e.g.$\delta_1 = 0.3, \delta_2 = 0.1$), a larger sample size is required to maintain the same power level as compared to when difference in the two treatments are bigger ($\delta_1 = 0.5, \delta_2 = 0.1$). For instance, when $\tau = 0.25, \delta_1 = 0.3, \delta_2 = 0.1$, we need 344 subjects to achieve 80% power, whereas, we only need 299 subjects to achieve the same power when $\delta_1 = 0.5, \delta_2 = 0.1$. When $\epsilon$ or $\tau$ increases, by fixing other parameters, smaller sample is needed. When the interim treatment effect is in the same direction as the final endpoints, smaller sample size is needed to maintain the same power level. For example, at interim selection, when $\nu_1^X - \nu_2^X > 0$, it implies the final endpoint be in the same direction if $\delta_1 > \delta_2$.

#### 3.4.2.2 Same Endpoints at Interim and Final Analyses

- Case I: $\delta_1 = \delta_2 = \delta$.

  When $\delta_1 = \delta_2 = \delta$, we can determine the sample size $n$ per group by solving

Table 3.5: Estimated sample size for $\delta_1 \neq \delta_2$ when $\nu_1^X - \nu_2^X = 0.1, \rho = 0.8, \sigma_X = \sigma_Y = 1, \alpha = 0.025$ with 80% or 90% power by $\tau, \delta_1, \delta_2$ and $\epsilon$ when different endpoints at interim and final analyses.

| | $\tau = 0.25, \rho = 0.8$( *i.e.* $\eta = 0.2$) | | | | | |
|---|---|---|---|---|---|---|
| | $\epsilon = 0$ | | $\epsilon = 0.1$ | | $\epsilon = 0.2$ | |
| $(\delta_1, \delta_2)$ | (0.3,0.1) | (0.5,0.1) | (0.3,0.1) | (0.5,0.1) | (0.3,0.1) | (0.5,0.1) |
| $1 - \beta = 0.8$ | 344 | 299 | 167 | 138 | 95 | 77 |
| $1 - \beta = 0.9$ | 615 | 613 | 258 | 252 | 138 | 131 |
| | $\tau = 0.5, \rho = 0.8$( *i.e.* $\eta = 0.28$) | | | | | |
| | $\epsilon = 0$ | | $\epsilon = 0.1$ | | $\epsilon = 0.2$ | |
| $(\delta_1, \delta_2)$ | (0.3,0.1) | (0.5,0.1) | (0.3,0.1) | (0.5,0.1) | (0.3,0.1) | (0.5,0.1) |
| $1 - \beta = 0.8$ | 275 | 186 | 147 | 107 | 88 | 66 |
| $1 - \beta = 0.9$ | 422 | 395 | 214 | 197 | 124 | 111 |
| | $\tau = 0.75, \rho = 0.8$( *i.e.* $\eta = 0.35$) | | | | | |
| | $\epsilon = 0$ | | $\epsilon = 0.1$ | | $\epsilon = 0.2$ | |
| $(\delta_1, \delta_2)$ | (0.3,0.1) | (0.5,0.1) | (0.3,0.1) | (0.5,0.1) | (0.3,0.1) | (0.5,0.1) |
| $1 - \beta = 0.8$ | 247 | 142 | 136 | 90 | 83 | 59 |
| $1 - \beta = 0.9$ | 352 | 290 | 190 | 162 | 114 | 97 |

Equation 3.15 given the $\alpha, \tau, c^*, \delta, \sigma_Y$, power $1 - \beta$ and $\epsilon$.

Table 3.6 provides an example of the estimated sample size for $\delta_1 = \delta_2 = \delta$ when $\sigma_Y = 1$ and $\alpha = 0.025$ with 80% or 90% power by $\tau, \delta$ and $\epsilon$. Given target treatment effect (i.e. $\delta$) and $\epsilon$, as $\tau$ increases, we need smaller sample size.

When the sum of $\delta$ and $\epsilon$ are the same, sample sizes are the same even individual $\delta$ and $\epsilon$ may vary. For example, in Table 3.6, at $\tau = 0.25$, when $\epsilon = 0, \delta = 0.2$, the estimated sample size needed to achieve 80% power is 378, which is the same as when $\epsilon = 0.1, \delta = 0.1$.

- Case II: $\delta_1 \neq \delta_2$.

  When $\delta_1 \neq \delta_2$, we can determine the sample size $n$ per group by solving Equation 3.16 given the $\alpha, \tau, c^*, \delta_1, \delta_2, \sigma_Y$, power $1 - \beta$ and $\epsilon$.

  Table 3.7 provides an example of the estimated sample size for $\delta_1 \neq \delta_2$ when $\sigma_Y = 1$ and $\alpha = 0.025$ with 80% or 90% power by $\tau, \delta_1, \delta_2$ and $\epsilon$. As $\tau$ or $\epsilon$ increases, the required sample size decreases.

Table 3.6: Estimated sample size for $\delta_1 = \delta_2 = \delta$ when $\sigma_Y = 1$ and $\alpha = 0.025$ with 80% or 90% power by $\tau, \delta, \delta$ and $\epsilon$ when same endpoints at interim and final analyses.

| | $\tau = 0.25$ | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\epsilon = 0$ | | | | | $\epsilon = 0.1$ | | | | | $\epsilon = 0.2$ | | | | |
| $\delta$ | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 |
| $1-\beta = 0.8$ | 1510 | 378 | 168 | 95 | 61 | 378 | 168 | 95 | 61 | 42 | 168 | 95 | 61 | 42 | 31 |
| $1-\beta = 0.9$ | 2020 | 505 | 225 | 127 | 81 | 505 | 225 | 127 | 81 | 57 | 225 | 127 | 81 | 57 | 42 |
| | $\tau = 0.5$ | | | | | | | | | | | | | | |
| | $\epsilon = 0$ | | | | | $\epsilon = 0.1$ | | | | | $\epsilon = 0.2$ | | | | |
| $\delta$ | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 |
| $1-\beta = 0.8$ | 1451 | 363 | 162 | 91 | 59 | 363 | 162 | 91 | 59 | 41 | 162 | 91 | 59 | 41 | 30 |
| $1-\beta = 0.9$ | 1940 | 485 | 216 | 122 | 78 | 485 | 216 | 122 | 78 | 54 | 216 | 122 | 78 | 54 | 40 |
| | $\tau = 0.75$ | | | | | | | | | | | | | | |
| | $\epsilon = 0$ | | | | | $\epsilon = 0.1$ | | | | | $\epsilon = 0.2$ | | | | |
| $\delta$ | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 |
| $1-\beta = 0.8$ | 1394 | 349 | 155 | 88 | 56 | 349 | 155 | 88 | 56 | 39 | 155 | 88 | 56 | 39 | 29 |
| $1-\beta = 0.9$ | 1860 | 465 | 207 | 117 | 75 | 465 | 207 | 117 | 75 | 52 | 207 | 117 | 75 | 52 | 38 |

Table 3.7: Estimated sample size for $\delta_1 \neq \delta_2$ when $\sigma_Y = 1$ and $\alpha = 0.025$ with 80% or 90% power by $\tau, \delta_1, \delta_2$ and $\epsilon$ when same endpoints at interim and final analyses.

| | $\tau = 0.25$ | | | | | |
|---|---|---|---|---|---|---|
| | $\epsilon = 0$ | | $\epsilon = 0.1$ | | $\epsilon = 0.2$ | |
| $(\delta_1, \delta_2)$ | (0.3,0.1) | (0.5,0.1) | (0.3,0.1) | (0.5,0.1) | (0.3,0.1) | (0.5,0.1) |
| $1-\beta = 0.8$ | 236 | 83 | 134 | 59 | 83 | 44 |
| $1-\beta = 0.9$ | 328 | 112 | 186 | 82 | 114 | 62 |
| | $\tau = 0.5$ | | | | | |
| | $\epsilon = 0$ | | $\epsilon = 0.1$ | | $\epsilon = 0.2$ | |
| $(\delta_1, \delta_2)$ | (0.3,0.1) | (0.5,0.1) | (0.3,0.1) | (0.5,0.1) | (0.3,0.1) | (0.5,0.1) |
| $1-\beta = 0.8$ | 212 | 76 | 120 | 53 | 76 | 40 |
| $1-\beta = 0.9$ | 281 | 99 | 160 | 71 | 102 | 53 |
| | $\tau = 0.75$ | | | | | |
| | $\epsilon = 0$ | | $\epsilon = 0.1$ | | $\epsilon = 0.2$ | |
| $(\delta_1, \delta_2)$ | (0.3,0.1) | (0.5,0.1) | (0.3,0.1) | (0.5,0.1) | (0.3,0.1) | (0.5,0.1) |
| $1-\beta = 0.8$ | 207 | 75 | 115 | 52 | 73 | 38 |
| $1-\beta = 0.9$ | 271 | 98 | 152 | 68 | 96 | 50 |

## 3.5 Normal Approximation

### 3.5.1 Density Function

Recall that the density function of the final test statistic $W^*$ is:

$$f_{W^*}(w) = pf_1(w - w_1 - \sqrt{\frac{n}{2\sigma_Y^2}}\epsilon) + qf_2(w - w_2 - \sqrt{\frac{n}{2\sigma_Y^2}}\epsilon).$$

Using results in Shun et al. (2008), the density function of $W^*$ can be approximated by:

$$f_{W^*}(w) \approx \frac{p}{\sigma_1}\phi(\frac{w - w_1 - \mu_1 - \sqrt{\frac{n}{2\sigma_Y^2}}\epsilon}{\sigma_1}) + \frac{q}{\sigma_2}\phi(\frac{w - w_2 - \mu_2 - \sqrt{\frac{n}{2\sigma_Y^2}}\epsilon}{\sigma_2}), \tag{3.17}$$

where

$$\begin{aligned} \mu_1 = \frac{\Lambda}{p}, \sigma_1^2 = 1 - \lambda\eta\mu_1 - \mu_1^2, \\ \mu_2 = \frac{\Lambda}{q}, \sigma_2^2 = 1 + \lambda\eta\mu_2 - \mu_2^2, \end{aligned} \tag{3.18}$$

with $\Lambda = \frac{\eta}{\sqrt{2\pi}}e^{-\frac{1}{2}\lambda^2}, p = \Phi(\lambda), q = 1 - p$; and on the other hand, when a surrogate endpoint is <u>not</u> used, $\lambda$ depends on $\eta, w_1$ and $w_2$, to be more specific, $\lambda = 2\eta(w_1 - w_2)$. We need only to replace $\lambda$ in (3.18), $\Lambda$ and $p$ with $\lambda = 2\eta(w_1 - w_2)$.

### 3.5.2 Type I Error

By Shun et al. (2008), under $H_0$, $\frac{W^* - \mu_0}{\sigma_0}$ can be approximated by a standard normal distribution, where

$$\mu_0 = \sqrt{\frac{2}{\pi}}\eta, \sigma_0^2 = 1 - \frac{2}{\pi}\eta^2. \tag{3.19}$$

Therefore, the rejection region $\Omega$ based on $c^*$, $\Omega = \{W^* : W^* > c^*\}$ can be approximated by a region $\tilde{\Omega}$, $\tilde{\Omega} = \{W^* : W^* > z_\alpha\sigma_0 + \mu_0\}$. When a surrogate endpoint is used, $\eta = \frac{\sqrt{\tau}}{2}\rho$; when a surrogate endpoint is not used, $\eta = \frac{\sqrt{\tau}}{2}$.

### 3.5.3 Power and Sample Size Estimation using Normal Approximation

#### 3.5.3.1 Different Endpoints at Interim and Final Analyses

When different endpoints are used at interim and final analyses, the power $1-\beta$ is calculated by

$$\begin{aligned}1-\beta =& \Pr(W^* > c^*; H_1) = \gamma(c^*; \lambda, w_1, w_2, \eta)\\ =& 1 - \int_{-\infty}^{c^*-w_1-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon} \Phi\left(\frac{\lambda+\eta z}{\sqrt{1-\eta^2}}\right)\phi(z)\,dz\\ &- \int_{-\infty}^{c^*-w_2-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon} \Phi\left(\frac{-\lambda+\eta z}{\sqrt{1-\eta^2}}\right)\phi(z)\,dz.\end{aligned}$$

- Case I: $\delta_1 \neq \delta_2$.

  When $\delta_1 \neq \delta_2$, the power $1-\beta$ can be calculated as:

$$\begin{aligned}1-\beta =& 1 - \int_{-\infty}^{c^*-\sqrt{\frac{n}{2\sigma_Y^2}}\delta_1-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon} \Phi\left(\frac{\lambda+\eta z}{\sqrt{1-\eta^2}}\right)\phi(z)\,dz\\ &- \int_{-\infty}^{c^*-\sqrt{\frac{n}{2\sigma_Y^2}}\delta_2-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon} \Phi\left(\frac{-\lambda+\eta z}{\sqrt{1-\eta^2}}\right)\phi(z)\,dz\\ =& 1 - pF_1(c^* - \sqrt{\frac{n}{2\sigma_Y^2}}\delta_1 - \sqrt{\frac{n}{2\sigma_Y^2}}\epsilon) - pF_2(c^* - \sqrt{\frac{n}{2\sigma_Y^2}}\delta_2 - \sqrt{\frac{n}{2\sigma_Y^2}}\epsilon)\\ =& 1 - p\beta_1 - q\beta_2,\end{aligned}$$

  where $\beta_j = F_j(c^* - \sqrt{\frac{n}{2\sigma_Y^2}}\delta_j - \sqrt{\frac{n}{2\sigma_Y^2}}\epsilon)$, for $j = 1, 2$.

  By normal approximation,

$$\begin{aligned}\beta_j &= \Phi(-z_{\beta_j})\\ &= F_j(c^* - \sqrt{\frac{n}{2\sigma_Y^2}}\delta_j - \sqrt{\frac{n}{2\sigma_Y^2}}\epsilon)\\ &\simeq \Phi\left(\frac{c^* - \sqrt{\frac{n}{2\sigma_Y^2}}\delta_j - \sqrt{\frac{n}{2\sigma_Y^2}}\epsilon - \mu_j}{\sigma_j}\right),\end{aligned}$$

Therefore, the power $1-\beta$ can be approximated by:

$$\begin{aligned}1-\beta &= 1-pF_1(c^*-\sqrt{\frac{n}{2\sigma_Y^2}}\delta_1-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon)-pF_2(c^*-\sqrt{\frac{n}{2\sigma_Y^2}}\delta_2-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon)\\ &\simeq 1-p\Phi\left(\frac{c^*-\sqrt{\frac{n}{2\sigma_Y^2}}\delta_1-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon-\mu_1}{\sigma_1}\right)-q\Phi\left(\frac{c^*-\sqrt{\frac{n}{2\sigma_Y^2}}\delta_2-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon-\mu_2}{\sigma_2}\right).\end{aligned}$$

Let $w_{\beta_j}=-(c^*-\sqrt{\frac{n}{2\sigma_Y^2}}\delta_j-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon)$, for $j=1,2$, then

$$-z_{\beta_j}\sigma_j \simeq -w_{\beta_j}-\mu_j,$$

which means

$$w_{\beta_j} \simeq z_{\beta_j}\sigma_j-\mu_j.$$

Under null hypothesis, the type I error of size $\alpha$ is calculated by

$$\begin{aligned}\alpha &= \Pr(W^*>c^*;H_0)\\ &= \gamma(c^*;\lambda=0,w_1=w_2=-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon,\eta)\\ &= \left[1-2\int_{-\infty}^{c^*}\Phi\left(\frac{\eta z}{\sqrt{1-\eta^2}}\right)\phi(z)\,dz\right]\\ &= 1-F_0(c^*),\end{aligned}$$

where $F_0(w)=2\int_{-\infty}^{w}\Phi\left(\frac{\eta z}{\sqrt{1-\eta^2}}\right)\phi(z)\,dz$.

By normal approximation,

$$1-\alpha=\Phi(z_\alpha)=F_0(c^*)\simeq\Phi\left(\frac{c^*-\mu_0}{\sigma_0}\right).$$

Hence,

$$z_\alpha\sigma_0 \simeq c^*-\mu_0,$$

we have

$$c^* \simeq z_\alpha\sigma_0+\mu_0.$$

Therefore,

$$c^* + w_{\beta_j} = \sqrt{\frac{n}{2\sigma_Y^2}}\,(\delta_j + \epsilon) \simeq z_\alpha \sigma_0 + \mu_0 + z_{\beta_j}\sigma_j - \mu_j, \text{ for } j = 1, 2.$$

The sample size $n$ per group needed for the study to maintain the power of $1-\beta$ can be approximated by

$$n \simeq 2\left(z_\alpha \sigma_0 + \mu_0 + z_{\beta_j}\sigma_j - \mu_j\right)^2 \left(\frac{\sigma_Y}{\delta_j + \epsilon}\right)^2, \text{ for } j = 1 \text{ or } 2, \tag{3.20}$$

subject to the constraints

$$\begin{cases} \frac{z_\alpha \sigma_0 + \mu_0 + z_{\beta_1}\sigma_1 - \mu_1}{z_\alpha \sigma_0 + \mu_0 + z_{\beta_2}\sigma_2 - \mu_2} = \frac{(\delta_1 + \epsilon)}{(\delta_2 + \epsilon)}, \\ 1 - \beta = 1 - p\beta_1 - q\beta_2. \end{cases} \tag{3.21}$$

In order to solve Equation 3.20 subject to constraints in (3.21), we discuss the following two scenarios:

1. Pre-determined timing of the interim selection $\tau$:

   When the timing of the interim selection is pre-determined: given $\tau$, $\rho, \nu_1^X - \nu_2^X$, and $\sigma_X$, we can calculate $\eta$, and $\lambda$ as a function of $n$:

$$\begin{aligned} \eta =& \frac{\sqrt{\tau}}{2}\rho, \\ \lambda =& \sqrt{\frac{n_1}{2\sigma_X^2}}\left(\nu_1^X - \nu_2^X\right) = \sqrt{\frac{n\tau}{2\sigma_X^2}}\left(\nu_1^X - \nu_2^X\right). \end{aligned} \tag{3.22}$$

   We can plug the parameters in (3.22) into (3.18) and (3.19) to get $\mu_j$ and $\sigma_j$, for $j = 0, 1, 2$. We also have

$$z_{\beta_j} = \frac{\sqrt{\frac{n}{2\sigma_Y^2}}\,(\delta_j + \epsilon) - z_\alpha \sigma_0 - \mu_0 + \mu_j}{\sigma_j}, \text{ for } j = 1, 2. \tag{3.23}$$

   Hence, given $\alpha,\ 1-\beta, \sigma_Y, \delta$'s and $\epsilon$, by solving the following three non-linear

Table 3.8: Comparison of sample size estimates using the exact distribution (upper row) and normal approximation (lower row in parenthesis) when $\tau = 0.5, \alpha = 0.025, \rho = 0.8, \nu_1^X - \nu_2^X = 0.1$ and $\sigma_Y = \sigma_X = 1$, for different endpoints at interim and final analyses.

| | $\epsilon = 0$ | | $\epsilon = 0.1$ | | $\epsilon = 0.2$ | |
|---|---|---|---|---|---|---|
| $(\delta_1, \delta_2)$ | (0.3,0.1) | (0.5,0.1) | (0.3,0.1) | (0.5,0.1) | (0.3,0.1) | (0.5,0.1) |
| $1 - \beta = 0.8$ | 275 | 88 | 147 | 58 | 88 | 41 |
| | (275) | (88) | (147) | (58) | (88) | (41) |
| $1 - \beta = 0.9$ | 422 | 124 | 214 | 80 | 124 | 56 |
| | (422) | (124) | (214) | (80) | (124) | (56) |

Note: The sample size estimates were rounded up to the integer.

equations, we can get solutions of the three unknown parameters $n$, $\beta_1$ and $\beta_2$:

$$\begin{cases} z_{\beta_1} = \frac{\sqrt{\frac{n}{2\sigma_Y^2}}(\delta_1+\epsilon) - z_\alpha \sigma_0 - \mu_0 + \mu_1}{\sigma_1}, \\ z_{\beta_2} = \frac{\sqrt{\frac{n}{2\sigma_Y^2}}(\delta_2+\epsilon) - z_\alpha \sigma_0 - \mu_0 + \mu_2}{\sigma_2}, \\ 1 - \beta = 1 - p\beta_1 - q\beta_2. \end{cases} \tag{3.24}$$

Mathematica program is provided in appendix D.1 to solve the above non-linear equations.

Table 3.8 provides an example to show the sample size estimates based on the exact distribution and based on the normal approximation approach. We pre-determined the timing of the interim selection to be $\tau = 0.5$. The numbers on the top show the sample size estimates from the exact distribution, and the numbers in the parenthesis is the sample size estimates from the normal approximation. We can see that the sample size estimates between these two approaches are very close.

2. Pre-determined winning probability $p$:

   During the planning stage of a two-stage winner design, if the winning probability $p$ is pre-determined: given $p, \rho, \sigma_X$, and $\nu_1^X - \nu_2^X$ we can calculate $n_1$ and $\lambda$, and $\eta$ as a function of $n$:

$$n_1 = \frac{2\sigma_X^2 {z_p}^2}{\left(\nu_1^X - \nu_2^X\right)^2},$$

Table 3.9: Comparison of sample size estimates using the exact distribution (upper row) and normal approximation (lower row in parenthesis) when $p = 0.65, \alpha = 0.025, \rho = 0.8, \nu_1^X - \nu_2^X = 0.1$ and $\sigma_Y = \sigma_X = 1$, for different endpoints at interim and final analyses.

| | $\epsilon = 0$ | $\epsilon = 0.1$ | $\epsilon = 0.2$ |
|---|---|---|---|
| $(\delta_1, \delta_2)$ | (0.15, 0.1) | (0.15, 0.1) | (0.15, 0.1) |
| $1 - \beta = 0.8$ | $n_1 = 30$ | $n_1 = 30$ | $n_1 = 30$ |
| | $(n_1 = 30)$ | $(n_1 = 30)$ | $(n_1 = 30)$ |
| n | 964 | 295 | 142 |
| | (964) | (295) | (142) |
| $1 - \beta = 0.9$ | $n_1 = 30$ | $n_1 = 30$ | $n_1 = 30$ |
| | $(n_1 = 30)$ | $(n_1 = 30)$ | $(n_1 = 30)$ |
| n | 1394 | 404 | 192 |
| | (1394) | (404) | (192) |

Note: The sample size estimates were rounded up to the integer.

$$\lambda = \sqrt{\frac{n_1}{2\sigma_X^2}} \left(\nu_1^X - \nu_2^X\right), \qquad \eta = \frac{\sqrt{\tau}}{2}\rho = \frac{\rho}{2}\sqrt{\frac{n_1}{n}}. \tag{3.25}$$

We can plug the parameters in (3.25) into (3.18) and (3.19) to get $\mu_j$ and $\sigma_j$, for $j = 0, 1, 2$. We also have Equation 3.23. Hence, given $\alpha, 1 - \beta, \sigma_Y, \delta$'s and $\epsilon$, by solving the following 3 non-linear equations in (3.24), we can get the solutions for the 3 unknown parameters $n$, $\beta_1$ and $\beta_2$. Mathematica program is provided in appendix D.2 to solve the non-linear equations.

Table 3.9 provides an example to show the sample size estimates based on the exact distribution and based on the normal approximation approach. We pre-determined the winning probability $p = 0.65$. The numbers on the top show the sample size estimates from the exact distribution, and the numbers in the parenthesis is the sample size estimates from the normal approximation. We can see that the sample size estimates between these two approaches are very close.

- Case II: $\delta_1 = \delta_2 = \delta$.

When $\delta_1 = \delta_2 = \delta$, it is a special case of Case I. Using the same normal approximation procedure described in Case I, we can get the sample size estimates.

Table 3.10: Comparison of sample size estimates using the exact distribution (upper row) and normal approximation (lower row in parenthesis) when $\tau = 0.25, \alpha = 0.025, \nu_1^X - \nu_2^X = 0.1, \rho = 0.8$, and $\sigma_Y = \sigma_X = 1$, for different endpoints at interim and final analyses.

| | $\epsilon = 0$ | | | | |
|---|---|---|---|---|---|
| $\delta$ | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 |
| $1-\beta = 0.8$ | 1658 | 394 | 173 | 97 | 62 |
| | (1657) | (394) | (173) | (97) | (62) |
| $1-\beta = 0.9$ | 2226 | 530 | 232 | 130 | 83 |
| | (2225) | (530) | (232) | (130) | (83) |
| | $\epsilon = 0.1$ | | | | |
| $\delta$ | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 |
| $1-\beta = 0.8$ | 394 | 173 | 97 | 62 | 43 |
| | (394) | (173) | (97) | (62) | (43) |
| $1-\beta = 0.9$ | 530 | 232 | 130 | 83 | 58 |
| | (530) | (232) | (130) | (83) | (58) |
| | $\epsilon = 0.2$ | | | | |
| $\delta$ | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 |
| $1-\beta = 0.8$ | 173 | 97 | 62 | 43 | 32 |
| | (173) | (97) | (62) | (43) | (32) |
| $1-\beta = 0.9$ | 232 | 130 | 83 | 58 | 42 |
| | (232) | (130) | (83) | (58) | (42) |

Note: The sample size estimates were rounded up to the integer.

Table 3.10 provides an example to show the comparison of the sample size estimates using Equation 3.12 based on the exact distribution and based on the normal approximation. We pre-determined the information time $\tau = 0.25$. The numbers on the top is the sample size estimates from the exact distribution, and the numbers in the parenthesis is the sample size from the normal approximation. We can see that the sample size estimates between these two approaches are very close.

Table 3.11 provides an example to show the sample size estimates based on the exact distribution and based on the normal approximation approach. We pre-determined the winning probability $p = 0.65$. The numbers on the top show the sample size estimations from the exact distribution, and the numbers in the parenthesis is the sample size estimates from the normal approximation. We can see that the sample size estimates between these two approaches are very close.

Table 3.11: Comparison of sample size estimates using the exact distribution (upper row) and normal approximation (lower row in parenthesis) when $p = 0.65, \alpha = 0.025, \rho = 0.8, \nu_1^X - \nu_2^X = 0.1$ and $\sigma_Y = \sigma_X = 1$, for different endpoints at interim and final analyses.

| | $\epsilon = 0$ | | $\epsilon = 0.1$ | | $\epsilon = 0.2$ | |
|---|---|---|---|---|---|---|
| $\delta$ | 0.1 | 0.2 | 0.1 | 0.2 | 0.1 | 0.2 |
| $1-\beta = 0.8$ | $n_1 = 30$ | $n_1 = 30$ | $n_1 = 30$ | $n_1 = 30$ | $n_1 = 30$ | $n_1 = 30$ |
| | $(n_1 = 30)$ | $(n_1 = 30)$ | $(n_1 = 30)$ | $(n_1 = 30)$ | $(n_1 = 30)$ | $(n_1 = 30)$ |
| n | 1571 | 392 | 392 | 173 | 173 | 97 |
| | (1571) | (392) | (392) | (173) | (173) | (97) |
| $1-\beta = 0.9$ | $n_1 = 30$ | $n_1 = 30$ | $n_1 = 30$ | $n_1 = 30$ | $n_1 = 30$ | $n_1 = 30$ |
| | $(n_1 = 30)$ | $(n_1 = 30)$ | $(n_1 = 30)$ | $(n_1 = 30)$ | $(n_1 = 30)$ | $(n_1 = 30)$ |
| n | 2103 | 525 | 525 | 232 | 232 | 130 |
| | (2103) | (525) | (525) | (232) | (232) | (130) |

Note: The sample size estimates were rounded up to the integer.

#### 3.5.3.2 Same Endpoints at Interim and Final Analyses

When the same endpoints are used at the interim and the final analyses, we can calculate the power of $1-\beta$ as:

$$\begin{aligned}
1-\beta &= \Pr(W^* > c^*; H_1) = \gamma(c^*; w_1, w_2, \eta) \\
&= 1 - \int_{-\infty}^{c^* - w_1 - \sqrt{\frac{n}{2\sigma_Y^2}}\epsilon} \Phi\left(\frac{2\eta(w_1 - w_2) + \eta z}{\sqrt{1-\eta^2}}\right) \phi(z)\, dz \\
&\quad - \int_{-\infty}^{c^* - w_2 - \sqrt{\frac{n}{2\sigma_Y^2}}\epsilon} \Phi\left(\frac{-2\eta(w_1 - w_2) + \eta z}{\sqrt{1-\eta^2}}\right) \phi(z)\, dz,
\end{aligned}$$

- Case I: $\delta_1 \neq \delta_2$.

  When $\delta_1 \neq \delta_2$, the power $1-\beta$ is:

$$\begin{aligned}
1-\beta &= 1 - \int_{-\infty}^{c^* - \sqrt{\frac{n}{2\sigma_Y^2}}\delta_1 - \sqrt{\frac{n}{2\sigma_Y^2}}\epsilon} \Phi\left(\frac{2\eta\,(w_1 - w_2) + \eta z}{\sqrt{1-\eta^2}}\right) \phi(z)\, dz \\
&\quad - \int_{-\infty}^{c^* - \sqrt{\frac{n}{2\sigma_Y^2}}\delta_2 - \sqrt{\frac{n}{2\sigma_Y^2}}\epsilon} \Phi\left(\frac{-2\eta\,(w_1 - w_2) + \eta z}{\sqrt{1-\eta^2}}\right) \phi(z)\, dz \\
&= 1 - pF_1(c^* - \sqrt{\frac{n}{2\sigma_Y^2}}\delta_1 - \sqrt{\frac{n}{2\sigma_Y^2}}\epsilon) - qF_2(c^* - \sqrt{\frac{n}{2\sigma_Y^2}}\delta_2 - \sqrt{\frac{n}{2\sigma_Y^2}}\epsilon) \\
&= 1 - p\beta_1 - q\beta_2,
\end{aligned}$$

where $\beta_j = F_j(c^* - \sqrt{\frac{n}{2\sigma_Y^2}}\delta_j - \sqrt{\frac{n}{2\sigma_Y^2}}\epsilon)$, for $j = 1, 2$. $F_j(w) = \int_{-\infty}^{w} f_j(t)dt$, with the density function $f_1(w) = \frac{1}{p}\Phi(\frac{2\eta(w_1-w_2)+\eta w}{\sqrt{1-\eta^2}})\phi(w)$, $f_2(w) = \frac{1}{q}\Phi(\frac{-2\eta(w_1-w_2)+\eta w}{\sqrt{1-\eta^2}})\phi(w)$.

Following the same normal approximation procedure as described in Section 3.5.3.1, the sample size $n$ per group needed for a study to have power of $1-\beta$ can be determined by Equation 3.20 subject to the constraints (3.21). In order to solve Equation 3.20 subject to constraints in (3.21), we discuss the following two scenarios:

1. Pre-determined timing of the interim selection $\tau$ :

   If the timing of the interim selection is pre-determined: given $\tau, \sigma_Y, \delta_1$ and $\delta_2$, we can calculate $\eta$, and $\lambda$ as a function of $n$:

$$\begin{aligned} \eta =& \frac{\sqrt{\tau}}{2}, \\ \lambda =& \sqrt{\frac{n\tau}{2\sigma_Y^2}}(\boldsymbol{\delta_1 - \delta_2}). \end{aligned} \tag{3.26}$$

   We can plug the parameters in (3.26) into (3.18) and (3.19) to get $\mu_j$ and $\sigma_j$, for $j = 0, 1, 2$.

   We also have Equation 3.23. Hence, given $\alpha, 1-\beta$ and $\epsilon$, by solving the 3 non-linear equations in (3.24), we can get solutions for the 3 unknown parameters $n$, $\beta_1$, and $\beta_2$. Mathematica program is provided in appendix D.3 to solve the non-linear equations.

   Table 3.12 provides an example to show the sample size estimates based on the exact distribution and based on the normal approximation approach. We pre-determined the timing of the interim selection to be $\tau = 0.5$. The numbers on the top show the sample size estimates from the exact distribution, and the numbers in the parenthesis is the sample size estimates from the normal approximation. We can see that the sample size estimates between these two approaches are very close.

2. Pre-determined winning probability $p$:

   During the planning stage of a two-stage winner design, if the winning probability $p$ is pre-determined: given $p, \delta_1, \delta_2$ and $\sigma_Y$, we can calculate $n_1$ and $\lambda$, and

Table 3.12: Comparison of sample size estimates using the exact distribution (upper row) and normal approximation (lower row in parenthesis) when $\tau = 0.5, \alpha = 0.025, \rho = 1$ and $\sigma_Y = 1$, for same endpoints at interim and final analyses.

| | $\epsilon = 0$ | | $\epsilon = 0.1$ | | $\epsilon = 0.2$ | |
|---|---|---|---|---|---|---|
| $(\delta_1, \delta_2)$ | (0.3,0.1) | (0.5,0.1) | (0.3,0.1) | (0.5,0.1) | (0.3,0.1) | (0.5,0.1) |
| $1 - \beta = 0.8$ | 212 | 76 | 120 | 53 | 76 | 40 |
| | (212) | (75) | (120) | (53) | (76) | (40) |
| $1 - \beta = 0.9$ | 281 | 99 | 160 | 71 | 102 | 53 |
| | (280) | (99) | (160) | (70) | (101) | (53) |

Note: The sample size estimates were rounded up to the integer.

$\eta$ as a function of $n$:

$$
\begin{aligned}
n_1 &= \frac{2\sigma_Y^2 {z_p}^2}{(\delta_1 - \delta_2)^2}, \\
\lambda &= \sqrt{\frac{n_1}{2\sigma_Y^2}}(\boldsymbol{\delta_1 - \delta_2}), \\
\eta &= \frac{\sqrt{\tau}}{2} = \frac{1}{2}\sqrt{\frac{n_1}{n}}.
\end{aligned} \tag{3.27}
$$

We can plug the parameters in (3.27) into (3.18) and (3.19) to get $\mu_j$ and $\sigma_j$, for $j = 0, 1, 2$. We also have Equation 3.23. Hence, given $\alpha, 1 - \beta$, and $\epsilon$, by solving the following 3 non-linear equations in (3.24), we can get solutions for the three unknown parameters $n$, $\beta_1$, and $\beta_2$. Mathematica program is provided in appendix D.4 to solve the non-linear equations.

Table 3.13 provides an example to show the sample size estimates based on the exact distribution and based on the normal approximation approach. We pre-determined the winning probability $p = 0.65$. The numbers on the top show the sample size estimations from the exact distribution, and the numbers in the parenthesis is the sample size estimates from the normal approximation. We can see that the sample size estimations between these two approaches are very close.

- Case II: $\delta_1 = \delta_2 = \delta$.

Table 3.13: Comparison of sample size estimates using the exact distribution (upper row) and normal approximation (lower row in parenthesis) when $p = 0.65, \alpha = 0.025, \rho = 1$ and $\sigma_Y = 1$, for same endpoints at interim and final analyses.

| | $\epsilon = 0$ | $\epsilon = 0.1$ | $\epsilon = 0.2$ |
|---|---|---|---|
| $(\delta_1, \delta_2)$ | (0.15, 0.1) | (0.15, 0.1) | (0.15, 0.1) |
| $1 - \beta = 0.8$ | $n_1 = 119$ | $n_1 = 119$ | $n_1 = 119$ |
| | $(n_1 = 119)$ | $(n_1 = 119)$ | $(n_1 = 119)$ |
| n | 946 | 280 | 126 |
| | (945) | (279) | (124) |
| $1 - \beta = 0.9$ | $n_1 = 119$ | $n_1 = 119$ | $n_1 = 119$ |
| | $(n_1 = 119)$ | $(n_1 = 119)$ | $(n_1 = 119)$ |
| n | 1367 | 386 | 175 |
| | (1367) | (385) | (174) |

Note: The sample size estimates were rounded up to the integer.

When $\delta_1 = \delta_2 = \delta$, the power $1 - \beta$ can be calculated by

$$\begin{aligned} 1 - \beta &= 1 - 2 \int_{-\infty}^{c^* - \sqrt{\frac{n}{2\sigma_Y^2}}\delta - \sqrt{\frac{n}{2\sigma_Y^2}}\epsilon} \Phi\left(\frac{\eta z}{\sqrt{1-\eta^2}}\right) \phi(z)\, dz \\ &= 1 - F_0(c^* - \sqrt{\frac{n}{2\sigma_Y^2}}\delta - \sqrt{\frac{n}{2\sigma_Y^2}}\epsilon), \end{aligned}$$

where $F_0$ denotes the distribution function with density function $f_0(w) = 2\Phi(\frac{\eta w}{\sqrt{1-\eta^2}})\phi(w)$, with mean $\mu_0 = \sqrt{\frac{2}{\pi}}\eta$ and variance $\sigma_0^2 = 1 - (\frac{2}{\pi})\eta^2$.

By Shun et al.(2008), the sample size for $\delta_1 = \delta_2 = \delta$ can be approximated by:

$$n \simeq 2\sigma_0^2 (z_\beta + z_\alpha)^2 (\frac{\sigma_Y}{\delta + \epsilon})^2, \tag{3.28}$$

where $\sigma_0^2 = 1 - (\frac{2}{\pi})\eta^2 = 1 - \frac{\tau}{2\pi}$.

From Equation 3.28, given $\tau, \delta, \alpha, \beta, \sigma_Y$ and $\epsilon$, we can get the sample size estimates. If there is a need to predetermine $n_1$ instead of $\tau$, we can replace $\tau$ by $\frac{n_1}{n}$ in Equation 3.28. The sample size can be determined by:

$$\begin{aligned} n &\simeq 2\sigma_0^2 (z_\beta + z_\alpha)^2 (\frac{\sigma_Y}{\delta + \epsilon})^2 \\ &= (2 - \frac{n_1}{n\pi})(z_\beta + z_\alpha)^2 (\frac{\sigma_Y}{\delta + \epsilon})^2. \end{aligned}$$

Solving the above equation, we have

$$n \simeq (z_\beta + z_\alpha)^2 (\frac{\sigma_Y}{\delta + \epsilon})^2 \left(1 + \sqrt{1 - \frac{n_1}{\pi (z_\beta + z_\alpha)^2 (\frac{\sigma_Y}{\delta+\epsilon})^2}}\right). \tag{3.29}$$

Notice that when no surrogate endpoints are used, Equations 3.28 and 3.29 for estimating the sample size for a non-inferiority trial as similar to the ones in Shun et al. (2008). The differences are: 1) we assume $\rho = 1$ because no surrogate endpoints are used; 2) sample size estimates depend on the non-inferiority margin $\epsilon$.

Table 3.14 provides an example to show the sample size estimates based on the exact distribution and based on the normal approximation approach. We pre-determined the timing of the interim selection at $\tau = 0.5$. The numbers on the top show the sample size estimates from the exact distribution, and the numbers in the parenthesis is the sample size estimates from the normal approximation. We can see that the sample size estimates between these two approaches are very close.

When $\delta_1 = \delta_2 = \delta$, the winning probability $p = \Phi(2\eta(w_1 - w_2)) = \Phi(0) = \frac{1}{2}$ always equals 0.5. Which means that there is no need to choose a winner at the interim analysis as there is equal chance to select either arm. In addition, the denominator of the $n_1 = \frac{2\sigma_Y^2 {z_p}^2}{(\delta_1 - \delta_2)^2}$ becomes 0 and undefined. Therefore, when $\delta_1 = \delta_2 = \delta$, it is meaningless to pre-determine $p$ for a two-stage winner design.

Table 3.15 and 3.16 summarized the difference among the sample size estimation formula using normal approximation with pre-determine $\tau$, and pre-determined $p$, respectively.

Table 3.14: Comparison of sample size estimates using the exact distribution (upper row) and normal approximation (lower row in parenthesis) when $\tau = 0.5, \alpha = 0.025, \rho = 1$ and $\sigma_Y = 1$, for same endpoints at interim and final analyses.

| | $\epsilon = 0$ | | | | |
|---|---|---|---|---|---|
| $\delta$ | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 |
| $1-\beta = 0.8$ | 1451 | 363 | 162 | 91 | 59 |
| | (1445) | (362) | (161) | (91) | (58) |
| $1-\beta = 0.9$ | 1940 | 485 | 216 | 122 | 78 |
| | (1935) | (484) | (215) | (121) | (78) |
| | $\epsilon = 0.1$ | | | | |
| $\delta$ | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 |
| $1-\beta = 0.8$ | 363 | 162 | 91 | 59 | 41 |
| | (362) | (161) | (91) | (58) | (41) |
| $1-\beta = 0.9$ | 485 | 216 | 122 | 78 | 54 |
| | (484) | (215) | (121) | (78) | (54) |
| | $\epsilon = 0.2$ | | | | |
| $\delta$ | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 |
| $1-\beta = 0.8$ | 162 | 91 | 59 | 41 | 30 |
| | (161) | (91) | (58) | (41) | (30) |
| $1-\beta = 0.9$ | 216 | 122 | 78 | 54 | 40 |
| | (215) | (121) | (78) | (54) | (40) |

Note: The sample size estimates were rounded up to the integer.

Table 3.15: Summary of sample size estimation formula using normal approximation with pre-determined $\tau$.

| | **Different Endpoints** |
|---|---|
| Given $\tau$, $\rho, \nu_1^X - \nu_2^X, \epsilon,$ $\sigma_X, \sigma_Y, \alpha, 1-\beta, \delta$'s | Case I: $\delta_1 \neq \delta_2$ |
| | $n \simeq 2(z_\alpha\sigma_0 + \mu_0 + z_{\beta_j}\sigma_j - \mu_j)^2(\frac{\sigma_Y}{\delta_j+\epsilon})^2$, for $j = 1$ or 2, subject to $\frac{z_\alpha\sigma_0+\mu_0+z_{\beta_1}\sigma_1-\mu_1}{z_\alpha\sigma_0+\mu_0+z_{\beta_2}\sigma_2-\mu_2} = \frac{(\delta_1+\epsilon)}{(\delta_2+\epsilon)}, 1-\beta = 1-p\beta_1-q\beta_2,$ where $\eta = \frac{\sqrt{\tau}}{2}\rho,$ $\lambda = \sqrt{\frac{n\tau}{2\sigma_X^2}}(\nu_1^X - \nu_2^X),$ |
| | Case II: $\delta_1 = \delta_2$ |
| | $n \simeq 2(z_\alpha\sigma_0 + \mu_0 + z_{\beta_j}\sigma_j - \mu_j)^2(\frac{\sigma_Y}{\delta+\epsilon})^2$, for $j = 1$ or 2, subject to $\frac{z_\alpha\sigma_0+\mu_0+z_{\beta_1}\sigma_1-\mu_1}{z_\alpha\sigma_0+\mu_0+z_{\beta_2}\sigma_2-\mu_2} = \frac{(\delta+\epsilon)}{(\delta+\epsilon)}, 1-\beta = 1-p\beta_1-q\beta_2,$ where $\eta = \frac{\sqrt{\tau}}{2}\rho,$ $\lambda = \sqrt{\frac{n\tau}{2\sigma_X^2}}(\nu_1^X - \nu_2^X),$ |
| | **Same Endpoints** |
| Given $\tau$, $\epsilon, \sigma_Y, \alpha,$ $1-\beta, \delta$'s | Case I: $\delta_1 \neq \delta_2$ |
| | $n \simeq 2(z_\alpha\sigma_0 + \mu_0 + z_{\beta_j}\sigma_j - \mu_j)^2(\frac{\sigma_Y}{\delta_j+\epsilon})^2$, for $j = 1$ or 2, subject to $\frac{z_\alpha\sigma_0+\mu_0+z_{\beta_1}\sigma_1-\mu_1}{z_\alpha\sigma_0+\mu_0+z_{\beta_2}\sigma_2-\mu_2} = \frac{(\delta_1+\epsilon)}{(\delta_2+\epsilon)}, 1-\beta = 1-p\beta_1-q\beta_2,$ where $\eta = \frac{\sqrt{\tau}}{2},$ $\lambda = \sqrt{\frac{n\tau}{2\sigma_Y^2}}(\delta_1 - \delta_2),$ |
| | Case II: $\delta_1 = \delta_2$ |
| | $n \simeq 2\sigma_0^2(z_\beta + z_\alpha)^2(\frac{\sigma_Y}{\delta+\epsilon})^2,$ where $\sigma_0^2 = 1 - (\frac{2}{\pi})\eta^2 = 1 - \frac{\tau}{2\pi},$ $\eta = \frac{\sqrt{\tau}}{2},$ |

Note: $\mu_1 = \frac{\Lambda}{p}, \sigma_1^2 = 1 - \lambda\eta\mu_1 - \mu_1^2, \mu_2 = \frac{\Lambda}{q}, \sigma_2^2 = 1 + \lambda\eta\mu_2 - \mu_2^2, \mu_0 = \sqrt{\frac{2}{\pi}}\eta, \sigma_0^2 = 1 - \frac{2}{\pi}\eta^2,$
$\Lambda = \frac{\eta}{\sqrt{2\pi}}e^{-\frac{1}{2}\lambda^2}, p = \Phi(\lambda), q = 1-p.$

Table 3.16: Summary of sample size estimation formula using normal approximation with pre-determined $p$.

| | **Different Endpoints** |
|---|---|
| Given $p$, $\rho, \nu_1^X - \nu_2^X, \epsilon, \alpha,$ $\sigma_X, \sigma_Y, 1-\beta, \delta$'s. | Case I: $\delta_1 \neq \delta_2$ |
| | $n \simeq 2(z_\alpha\sigma_0 + \mu_0 + z_{\beta_j}\sigma_j - \mu_j)^2(\frac{\sigma_Y}{\delta_j+\epsilon})^2$, for $j = 1$ or 2, subject to $\frac{z_\alpha\sigma_0+\mu_0+z_{\beta_1}\sigma_1-\mu_1}{z_\alpha\sigma_0+\mu_0+z_{\beta_2}\sigma_2-\mu_2} = \frac{(\delta_1+\epsilon)}{(\delta_2+\epsilon)}, 1-\beta = 1-p\beta_1-q\beta_2,$ with $n_1 = \frac{2\sigma_X^2 {z_p}^2}{(\nu_1^X-\nu_2^X)^2}$, $\lambda = \sqrt{\frac{n_1}{2\sigma_X^2}}(\nu_1^X - \nu_2^X)$, $\eta = \frac{\sqrt{\tau}}{2}\rho = \frac{\rho}{2}\sqrt{\frac{n_1}{n}}$. |
| | Case II: $\delta_1 = \delta_2$ |
| | $n \simeq 2(z_\alpha\sigma_0 + \mu_0 + z_{\beta_j}\sigma_j - \mu_j)^2(\frac{\sigma_Y}{\delta+\epsilon})^2$, for $j = 1$ or 2, subject to $\frac{z_\alpha\sigma_0+\mu_0+z_{\beta_1}\sigma_1-\mu_1}{z_\alpha\sigma_0+\mu_0+z_{\beta_2}\sigma_2-\mu_2} = \frac{(\delta+\epsilon)}{(\delta+\epsilon)}, 1-\beta = 1-p\beta_1-q\beta_2,$ with $n_1 = \frac{2\sigma_X^2 {z_p}^2}{(\nu_1^X-\nu_2^X)^2}$, $\lambda = \sqrt{\frac{n_1}{2\sigma_X^2}}(\nu_1^X - \nu_2^X)$, $\eta = \frac{\sqrt{\tau}}{2}\rho = \frac{\rho}{2}\sqrt{\frac{n_1}{n}}$. |
| | **Same Endpoints** |
| Given $p$, $\epsilon, \sigma_Y, \alpha,$ $1-\beta, \delta$'s | Case I: $\delta_1 \neq \delta_2$ |
| | $n \simeq 2(z_\alpha\sigma_0 + \mu_0 + z_{\beta_j}\sigma_j - \mu_j)^2(\frac{\sigma_Y}{\delta_j+\epsilon})^2$, for $j = 1$ or 2, subject to $\frac{z_\alpha\sigma_0+\mu_0+z_{\beta_1}\sigma_1-\mu_1}{z_\alpha\sigma_0+\mu_0+z_{\beta_2}\sigma_2-\mu_2} = \frac{(\delta_1+\epsilon)}{(\delta_2+\epsilon)}, 1-\beta = 1-p\beta_1-q\beta_2,$ with $n_1 = \frac{2\sigma_Y^2 {z_p}^2}{(\delta_1-\delta_2)^2}$, $\lambda = \sqrt{\frac{n_1}{2\sigma_Y^2}}(\delta_1 - \delta_2)$, $\eta = \frac{\sqrt{\tau}}{2} = \frac{1}{2}\sqrt{\frac{n_1}{n}}$. |
| | Case II: $\delta_1 = \delta_2$ |
| | NA |

Note: $\mu_1 = \frac{\Lambda}{p}, \sigma_1^2 = 1 - \lambda\eta\mu_1 - \mu_1^2, \mu_2 = \frac{\Lambda}{q}, \sigma_2^2 = 1 + \lambda\eta\mu_2 - \mu_2^2$, $\mu_0 = \sqrt{\frac{2}{\pi}}\eta, \sigma_0^2 = 1 - \frac{2}{\pi}\eta^2$, $\Lambda = \frac{\eta}{\sqrt{2\pi}}e^{-\frac{1}{2}\lambda^2}, p = \Phi(\lambda), q = 1-p.$

# Chapter 4

# Two-Stage Winner Design-Extension to the Trial with Three Experimental Arms

In Chapter 3, we developed the methodology for a non-inferiority trial in a two-stage winner design with two experimental treatment and one active control arms. In this chapter, we will extend the work to a two-stage winner design with three experimental treatment and one control arms. Similar to the setting in Chapter 3, a balanced sample size $n$ for each arm is planned and one interim analysis will be performed at the information time $\tau = \frac{n_1}{n}$, where $n_1$ $(< n)$ is the interim sample size. Let $\{X_i^{(j)} | i = 1, ..., n_1\}$ be the interim continuous measurements assumed to be independent identically distributed with $N(\nu_j^X, \sigma_X^2)$, where $j = 0$ denote the control arm and $j = 1, 2, 3$, denote the three experimental treatment arms. Let $\{Y_i^{(j)} | i = 1, ..., n\}, j = 0, 1, 2, 3$ denote the final continuous measurements that are assumed to be independent identically distributed with $N(\mu_j^Y, \sigma_Y^2)$. Assume the variances $\sigma_X^2$ and $\sigma_Y^2$ are known, and $X_i^{(j)}$ and $Y_i^{(j)}$ are correlated with a correlation $\rho$, i.e., $corr(X_i^{(j)}, Y_i^{(j)}) = \rho$.

In this chapter, we discuss two kinds of hypotheses, superiority and non-inferiority hypotheses, and we discuss two commonly used scenarios depends on whether the same or different endpoints are used at the interim and the final analyses.

## 4.1 Settings and Test Statistic

Let the interim sample means be $\bar{X}_{n_1}^{(j)} = (1/n_1) \sum_{i=1}^{n_1} X_i^{(j)}, j = 0, 1, 2, 3$ and the final sample means be $\bar{Y}_n^{(j)} = (1/n) \sum_{i=1}^{n} Y_i^{(j)}, j = 0, 1, 2, 3$. Similar to Shun et al. (2008), we use the

following procedure at the interim look:

$$\begin{cases} \text{Keep Treatment 1,} & \text{when } \bar{X}_{n_1}^{(1)} > \bar{X}_{n_1}^{(2)}, \text{and} \bar{X}_{n_1}^{(1)} > \bar{X}_{n_1}^{(3)}, \\ \text{Keep Treatment 2,} & \text{when } \bar{X}_{n_1}^{(2)} > \bar{X}_{n_1}^{(1)}, \text{and} \bar{X}_{n_1}^{(2)} > \bar{X}_{n_1}^{(3)}, \\ \text{Keep Treatment 3,} & \text{when } \bar{X}_{n_1}^{(3)} > \bar{X}_{n_1}^{(1)}, \text{and} \bar{X}_{n_1}^{(3)} > \bar{X}_{n_1}^{(2)}, \end{cases}$$

The goal of the design is to select only one experimental treatment at the interim analysis.

### 4.1.1 Test Statistic for Superiority Hypothesis

Let $\Delta_j = \mu_j^Y - \mu_0^Y$ be the unknown treatment effect of the jth experimental treatment versus active control at the final analysis, where $j = 1$ or 2 or 3. We consider the following hypotheses:

$$H_0 : \Delta_1 \leq 0 \text{ and } \Delta_2 \leq 0 \text{ and } \Delta_3 \leq 0$$

versus

$$H_1 : \Delta_1 > 0 \text{ or } \Delta_2 > 0, \text{ or } \Delta_3 > 0.$$

Assume the targeted treatment effects of the designs are $\Delta_j = \delta_j, j = 1, 2, 3$. Let the test statistics be

$$Z_n^{(j)} = \sqrt{\frac{n}{2\sigma_Y^2}}(\bar{Y}_n^{(j)} - \bar{Y}_n^{(0)}),$$

for j=1,2, and 3, respectively. Let

$$V_{n_1}^{(12)} = \sqrt{\frac{n_1}{2\sigma_X^2}}(\bar{X}_{n_1}^{(1)} - \bar{X}_{n_1}^{(2)}),$$

$$V_{n_1}^{(13)} = \sqrt{\frac{n_1}{2\sigma_X^2}}(\bar{X}_{n_1}^{(1)} - \bar{X}_{n_1}^{(3)}),$$

$$V_{n_1}^{(23)} = \sqrt{\frac{n_1}{2\sigma_X^2}}(\bar{X}_{n_1}^{(2)} - \bar{X}_{n_1}^{(3)}) = V_{n_1}^{(13)} - V_{n_1}^{(12)}.$$

It can be seen that

$$\eta = Cov(Z_n^{(1)}, V_{n_1}^{(12)}) = Cov(Z_n^{(1)}, V_{n_1}^{(13)}) = \frac{\sqrt{\tau}}{2}\rho,$$

$$\eta = -Cov(Z_n^{(2)}, V_{n_1}^{(12)}) = Cov(Z_n^{(2)}, V_{n_1}^{(23)}) = \frac{\sqrt{\tau}}{2}\rho,$$

Under $H_0$, $V_{n_1}^{(12)} \sim N(0,1), V_{n_1}^{(13)} \sim N(0,1),$ and $V_{n_1}^{(23)} \sim N(0,1)$. Under $H_1$, $V_{n_1}^{(12)} \sim N(\lambda^{(12)},1), V_{n_1}^{(13)} \sim N(\lambda^{(13)},1),$ and $V_{n_1}^{(23)} \sim N(\lambda^{(23)},1)$. where

$$\lambda^{(12)} = \sqrt{\frac{n_1}{2\sigma_X^2}}(\nu_1^X - \nu_2^X),$$

$$\lambda^{(13)} = \sqrt{\frac{n_1}{2\sigma_X^2}}(\nu_1^X - \nu_3^X),$$

$$\lambda^{(23)} = \lambda^{(13)} - \lambda^{(12)} = \sqrt{\frac{n_1}{2\sigma_X^2}}(\nu_2^X - \nu_3^X).$$

We define the final test statistic $W$ as follows:

$$W = \begin{cases} \frac{\bar{Y}_n^{(1)} - \bar{Y}_n^{(0)}}{\sqrt{\frac{2\sigma_Y^2}{n}}} = Z_n^{(1)}, & \text{if } \bar{X}_{n_1}^{(1)} = max(\bar{X}_{n_1}^{(1)}, \bar{X}_{n_1}^{(2)}, \bar{X}_{n_1}^{(3)}), \\ \frac{\bar{Y}_n^{(2)} - \bar{Y}_n^{(0)}}{\sqrt{\frac{2\sigma_Y^2}{n}}} = Z_n^{(2)}, & \text{if } \bar{X}_{n_1}^{(2)} = max(\bar{X}_{n_1}^{(1)}, \bar{X}_{n_1}^{(2)}, \bar{X}_{n_1}^{(3)}), \\ \frac{\bar{Y}_n^{(3)} - \bar{Y}_n^{(0)}}{\sqrt{\frac{2\sigma_Y^2}{n}}} = Z_n^{(3)}, & \text{if } \bar{X}_{n_1}^{(3)} = max(\bar{X}_{n_1}^{(1)}, \bar{X}_{n_1}^{(2)}, \bar{X}_{n_1}^{(3)}), \end{cases} \tag{4.1}$$

$H_0$ is rejected if $W > c$, where $c$ is the critical value to be determined in Section 4.3. The distribution of $W$ is introduced in Section 4.2 with the mathematical details provided in Appendix C.1.

### 4.1.2 Test Statistic for Non-inferiority Hypothesis

Use the same setting and notation as in the previous section. Let $\Delta_j = \mu_j^Y - \mu_0^Y$ be the unknown treatment effect of the jth experimental treatment versus the control at the final analysis, where $j = 1$ or 2 or 3. We consider the following non-inferiority hypotheses:

$$H_0 : \Delta_1 \leq -\epsilon \text{ and } \Delta_2 \leq -\epsilon \text{ and } \Delta_3 \leq -\epsilon$$

versus

$$H_1 : \Delta_1 > -\epsilon \text{ or } \Delta_2 > -\epsilon, \text{ or } \Delta_3 > -\epsilon$$

where $\epsilon$ ($> 0$) is the pre-specified non-inferiority margin which can be interpreted as the largest clinically and statistically acceptable difference. The alternative hypothesis $H_1$ states that either treatment 1 or treatment 2 or treatment 3 is non-inferior to the active control within an acceptable non-inferiority margin. To test the non-inferiority hypotheses with a pre-specified non-inferiority margin $\epsilon$ in the final analysis, we define the final test statistic $W^*$ as follows:

$$W^* = \begin{cases} \frac{\bar{Y}_n^{(1)} - \bar{Y}_n^{(0)} + \epsilon}{\sqrt{\frac{2\sigma_Y^2}{n}}} = Z_n^{(1)} + \sqrt{\frac{n}{2\sigma_Y^2}}\epsilon, & \text{if } \bar{X}_{n_1}^{(1)} = max(\bar{X}_{n_1}^{(1)}, \bar{X}_{n_1}^{(2)}, \bar{X}_{n_1}^{(3)}), \\ \frac{\bar{Y}_n^{(2)} - \bar{Y}_n^{(0)} + \epsilon}{\sqrt{\frac{2\sigma_Y^2}{n}}} = Z_n^{(2)} + \sqrt{\frac{n}{2\sigma_Y^2}}\epsilon, & \text{if } \bar{X}_{n_1}^{(2)} = max(\bar{X}_{n_1}^{(1)}, \bar{X}_{n_1}^{(2)}, \bar{X}_{n_1}^{(3)}), \\ \frac{\bar{Y}_n^{(3)} - \bar{Y}_n^{(0)} + \epsilon}{\sqrt{\frac{2\sigma_Y^2}{n}}} = Z_n^{(3)} + \sqrt{\frac{n}{2\sigma_Y^2}}\epsilon, & \text{if } \bar{X}_{n_1}^{(3)} = max(\bar{X}_{n_1}^{(1)}, \bar{X}_{n_1}^{(2)}, \bar{X}_{n_1}^{(3)}), \end{cases} \tag{4.2}$$

$H_0$ is rejected if $W^* > c^*$, where $c^*$ is the critical value to be determined in Section 4.3.

The distribution of W for superiority hypothesis is a special case of the distribution of $W^*$ for non-inferiority hypothesis. Therefore, we will focus on the distribution of $W^*$. In addition, the distribution of the test statistic and related statistical properties can be discuss in two scenarios depends on whether surrogate endpoint is used at the interim analysis or not. The first scenario is when surrogate endpoint is used at the interim analysis, in other words, different endpoints are used at the interim and the final analysis. The second scenario is when same endpoints are used at the interim and final analysis.

## 4.2 Distribution of the Test Statistic

### 4.2.1 Different Endpoints at Interim and Final Analyses

In Appendix C.1, we derived the distribution of the test statistic $W$ and have proved the following lemma:

**Lemma 4.2.1.**

$$F_W(w) = \sqrt{\frac{3}{4} - \eta^2} \int_{-\infty}^{w-w_1} \phi(z) \int_{-\infty}^{\frac{\lambda^{(12)} + \eta z}{\sqrt{1-\eta^2}}} \phi(s)\Phi\left(\frac{\left(\lambda^{(13)} + \eta z\right)\sqrt{1-\eta^2} - (\frac{1}{2} - \eta^2)s}{\sqrt{\frac{3}{4} - \eta^2}}\right) ds dz$$

$$+\sqrt{\frac{3}{4}-\eta^2}\int_{-\infty}^{w-w_2}\phi(z)\int_{-\infty}^{\frac{-\lambda^{(12)}+\eta z}{\sqrt{1-\eta^2}}}\phi(s)\Phi(\frac{\left(\lambda^{(13)}-\lambda^{(12)}+\eta z\right)\sqrt{1-\eta^2}-(\frac{1}{2}-\eta^2)s}{\sqrt{\frac{3}{4}-\eta^2}})dsdz$$

$$+\sqrt{\frac{3}{4}-\eta^2}\int_{-\infty}^{w-w_3}\phi(z)\int_{-\infty}^{\frac{-\lambda^{(13)}+\eta z}{\sqrt{1-\eta^2}}}\phi(s)\Phi(\frac{\left(\lambda^{(12)}-\lambda^{(13)}+\eta z\right)\sqrt{1-\eta^2}-(\frac{1}{2}-\eta^2)s}{\sqrt{\frac{3}{4}-\eta^2}})dsdz.$$

Therefore, the distribution of the test statistic $W^*$ for non-inferiority hypothesis can be derived as:

$$\begin{aligned}
F_{W^*}(w) =& \Pr\left(W^* < w, \lambda^{(12)}, \lambda^{(13)}, w_1, w_2, w_3, \eta\right)\\
=&\sqrt{\frac{3}{4}-\eta^2}\int_{-\infty}^{w-w_1-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon}\phi(z)\int_{-\infty}^{\frac{\lambda^{(12)}+\eta z}{\sqrt{1-\eta^2}}}\phi(s)\\
&\times\Phi(\frac{\left(\lambda^{(13)}+\eta z\right)\sqrt{1-\eta^2}-(\frac{1}{2}-\eta^2)s}{\sqrt{\frac{3}{4}-\eta^2}})dsdz\\
&+\sqrt{\frac{3}{4}-\eta^2}\int_{-\infty}^{w-w_2-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon}\phi(z)\int_{-\infty}^{\frac{-\lambda^{(12)}+\eta z}{\sqrt{1-\eta^2}}}\phi(s)\\
&\times\Phi(\frac{\left(\lambda^{(13)}-\lambda^{(12)}+\eta z\right)\sqrt{1-\eta^2}-(\frac{1}{2}-\eta^2)s}{\sqrt{\frac{3}{4}-\eta^2}})dsdz\\
&+\sqrt{\frac{3}{4}-\eta^2}\int_{-\infty}^{w-w_3-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon}\phi(z)\int_{-\infty}^{\frac{-\lambda^{(13)}+\eta z}{\sqrt{1-\eta^2}}}\phi(s)\\
&\times\Phi(\frac{\left(\lambda^{(12)}-\lambda^{(13)}+\eta z\right)\sqrt{1-\eta^2}-(\frac{1}{2}-\eta^2)s}{\sqrt{\frac{3}{4}-\eta^2}})dsdz. \qquad (4.3)
\end{aligned}$$

where

$$w_1=\sqrt{\frac{n}{2\sigma_Y^2}}\delta_1, w_2=\sqrt{\frac{n}{2\sigma_Y^2}}\delta_2, \text{ and } w_3=\sqrt{\frac{n}{2\sigma_Y^2}}\delta_3.$$

$\Phi(.)$ and $\phi(.)$ denote the C.D.F. and the P.D.F. of the standard normal distribution, respectively. Mathematical derivation can be found in Appendix C.1. When $\epsilon = 0$ in (4.3), it becomes the distribution of W for a two-stage winner design with superiority hypothesis as shown in Lemma 4.2.1.

### 4.2.2 Same Endpoints at Interim and Final Analyses

When the interim and the final endpoints are the same, then $\nu_j^X = \mu_j^Y$ for $j = 1, 2, 3$, and $\sigma_Y = \sigma_X$. Therefore, $\eta = \frac{\sqrt{\tau}}{2}\rho = \frac{\sqrt{\tau}}{2}$. The parameters $\lambda^{(12)}$ or $\lambda^{(13)}$ become the function of $\eta$ (or $\tau$), $w_1$ and $w_2$.

$$\begin{aligned}\lambda^{(12)} =& \sqrt{\frac{n_1}{2\sigma_X^2}}(\nu_1^X - \nu_2^X) = \sqrt{\frac{n_1}{2\sigma_X^2}}(\mu_1^Y - \mu_2^Y) = \sqrt{\frac{n_1}{2\sigma_X^2}}(\delta_1 - \delta_2)\\ =& \sqrt{\frac{n_1}{2\sigma_X^2}}\sqrt{\frac{2\sigma_Y^2}{n}}(w_1 - w_2) = \sqrt{\tau}(w_1 - w_2) = 2\eta(w_1 - w_2),\end{aligned}$$

$$\begin{aligned}\lambda^{(13)} =& \sqrt{\frac{n_1}{2\sigma_X^2}}(\nu_1^X - \nu_3^X) = \sqrt{\frac{n_1}{2\sigma_X^2}}(\mu_1^Y - \mu_3^Y) = \sqrt{\frac{n_1}{2\sigma_X^2}}(\delta_1 - \delta_3)\\ =& \sqrt{\frac{n_1}{2\sigma_X^2}}\sqrt{\frac{2\sigma_Y^2}{n}}(w_1 - w_3) = \sqrt{\tau}(w_1 - w_3) = 2\eta(w_1 - w_3).\end{aligned}$$

The distribution of $W^*$ becomes:

$$\begin{aligned}F_{W^*}(w) =& \Pr\left(W^* < w, \lambda^{(12)}, \lambda^{(13)}, w_1, w_2, w_3, \eta\right)\\ =& \sqrt{\frac{3}{4} - \eta^2}\int_{-\infty}^{w - w_1 - \sqrt{\frac{n}{2\sigma_Y^2}}\epsilon}\phi(z)\int_{-\infty}^{\frac{2\eta(w_1 - w_2) + \eta z}{\sqrt{1-\eta^2}}}\phi(s)\\ &\times \Phi\left(\frac{(2\eta(w_1 - w_3) + \eta z)\sqrt{1-\eta^2} - (\frac{1}{2} - \eta^2)s}{\sqrt{\frac{3}{4} - \eta^2}}\right)dsdz\\ &+\sqrt{\frac{3}{4} - \eta^2}\int_{-\infty}^{w - w_2 - \sqrt{\frac{n}{2\sigma_Y^2}}\epsilon}\phi(z)\int_{-\infty}^{\frac{2\eta(w_2 - w_1) + \eta z}{\sqrt{1-\eta^2}}}\phi(s)\\ &\times \Phi\left(\frac{(2\eta(w_2 - w_3) + \eta z)\sqrt{1-\eta^2} - (\frac{1}{2} - \eta^2)s}{\sqrt{\frac{3}{4} - \eta^2}}\right)dsdz\\ &+\sqrt{\frac{3}{4} - \eta^2}\int_{-\infty}^{w - w_3 - \sqrt{\frac{n}{2\sigma_Y^2}}\epsilon}\phi(z)\int_{-\infty}^{\frac{2\eta(w_3 - w_1) + \eta z}{\sqrt{1-\eta^2}}}\phi(s)\\ &\times \Phi\left(\frac{(2\eta(w_3 - w_2) + \eta z)\sqrt{1-\eta^2} - (\frac{1}{2} - \eta^2)s}{\sqrt{\frac{3}{4} - \eta^2}}\right)dsdz. \qquad (4.4)\end{aligned}$$

where

$$w_1 = \sqrt{\frac{n}{2\sigma_Y^2}}\delta_1, w_2 = \sqrt{\frac{n}{2\sigma_Y^2}}\delta_2, \text{ and } w_3 = \sqrt{\frac{n}{2\sigma_Y^2}}\delta_3.$$

When $\epsilon = 0$, it becomes the distribution of $W$ for a two-stage winner design with superiority hypothesis. Note the distribution of $W^*$ in (4.4) is in the same form as in (4.3) if we let $2\eta(w_1 - w_2) = \lambda^{(12)},\ 2\eta(w_1 - w_3) = \lambda^{(13)}$ and $2\eta(w_3 - w_2) = \lambda^{(12)} - \lambda^{(13)}$.

## 4.3 Type I Error

### 4.3.1 Different Endpoints at Interim and Final Analyses

#### 4.3.1.1 Non-inferiority Hypothesis

Given $\epsilon$, the type I error of size $\alpha$ test associated with $W^*$ is,

$$\begin{aligned}\alpha &= \sup_{\lambda^{(12)},\lambda^{(13)},\delta_1\le-\epsilon,\delta_2\le-\epsilon,\delta_3\le-\epsilon} Pr(W^* > c^*; \lambda^{(12)}, \lambda^{(13)}, \delta_1, \delta_2, \delta_3, \eta) \\ &= \sup_{\lambda^{(12)},\lambda^{(13)},w_1\le-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon,w_2\le-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon,w_3\le-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon} Pr(W^* > c^*; \lambda^{(12)}, \lambda^{(13)}, w_1, w_2, w_3, \eta),\end{aligned} \tag{4.5}$$

where $w_j = \sqrt{\frac{n}{2\sigma_Y^2}}\delta_j, j = 1, 2, 3$, and $c^*$ is the critical value. To determine $c^*$, we define a function $\gamma$ as:

$$\begin{aligned}&\gamma\left(b; \lambda^{(12)}, \lambda^{(13)}, w_1, w_2, w_3, \eta\right) = \Pr\left(W^* > b; \lambda^{(12)}, \lambda^{(13)}, w_1, w_2, w_3, \eta\right) \\ =&1 - \sqrt{\frac{3}{4} - \eta^2} \int_{-\infty}^{b - w_1 - \sqrt{\frac{n}{2\sigma_Y^2}}\epsilon} \phi(z) \int_{-\infty}^{\frac{\lambda^{(12)} + \eta z}{\sqrt{1-\eta^2}}} \phi(s) \\ &\times \Phi\left(\frac{\left(\lambda^{(13)} + \eta z\right)\sqrt{1-\eta^2} - (\frac{1}{2} - \eta^2)s}{\sqrt{\frac{3}{4} - \eta^2}}\right) ds dz \\ &- \sqrt{\frac{3}{4} - \eta^2} \int_{-\infty}^{b - w_2 - \sqrt{\frac{n}{2\sigma_Y^2}}\epsilon} \phi(z) \int_{-\infty}^{\frac{-\lambda^{(12)} + \eta z}{\sqrt{1-\eta^2}}} \phi(s) \\ &\times \Phi\left(\frac{\left(\lambda^{(13)} - \lambda^{(12)} + \eta z\right)\sqrt{1-\eta^2} - (\frac{1}{2} - \eta^2)s}{\sqrt{\frac{3}{4} - \eta^2}}\right) ds dz \\ &- \sqrt{\frac{3}{4} - \eta^2} \int_{-\infty}^{b - w_3 - \sqrt{\frac{n}{2\sigma_Y^2}}\epsilon} \phi(z) \int_{-\infty}^{\frac{-\lambda^{(13)} + \eta z}{\sqrt{1-\eta^2}}} \phi(s) \\ &\times \Phi\left(\frac{\left(\lambda^{(12)} - \lambda^{(13)} + \eta z\right)\sqrt{1-\eta^2} - (\frac{1}{2} - \eta^2)s}{\sqrt{\frac{3}{4} - \eta^2}}\right) ds dz,\end{aligned}$$

where $b$ is some cut-off value. The type I error rate in Equation 4.5 can be rewritten to:

$$\begin{aligned}\alpha &= \sup_{\lambda^{(12)},\lambda^{(13)},w_1\le-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon,w_2\le-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon,w_3\le-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon} Pr(W^* > c^*;\lambda^{(12)},\lambda^{(13)},w_1,w_2,w_3,\eta)\\ &= \sup_{\lambda^{(12)},\lambda^{(13)},w_1\le-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon,w_2\le-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon,w_3\le-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon} \gamma(c^*;\lambda^{(12)},\lambda^{(13)},w_1,w_2,w_3,\eta).\end{aligned}$$

In order to find the supremum of $\gamma(b;\lambda^{(12)},\lambda^{(13)},w_1,w_2,w_3,\eta)$, we have proven the following lemmas in Appendices C.2 and C.3. Notice that for superiority hypotheses ($\epsilon = 0$), if we use the same type of null hypothesis as in Shun et al. (2008),

$$H_0 : \Delta_1 = 0 \text{ and } \Delta_2 = 0 \text{ and } \Delta_3 = 0,$$

we don't need to use the following lemmas, and the generalization of the results in Shun et al. (2008) would be much easier. But we prefer to consider the superiority test as a special case of the framework laid out in this section.

**Lemma 4.3.1.** *Given $b,\eta$, $\lambda^{(12)}$ and $\lambda^{(13)}$, $\gamma(b;\lambda^{(12)},\lambda^{(13)},w_1,w_2,w_3,\eta)$ is monotonically increasing with $w_1,w_2$ and $w_3$.*

**Lemma 4.3.2.** *Given $b,\eta$, $w_1,w_2$ and $w_3$, $\lambda^{(12)} = \lambda^{(13)} = 0$ is a critical point of $\gamma(b;\lambda^{(12)},\lambda^{(13)},w_1,w_2,w_3,\eta)$.*

The uniqueness of the critical point in Lemma 4.3.2 would imply that the global maximum of $\gamma$ is at $\lambda^{(12)} = \lambda^{(13)} = 0$, but it is hard to theoretically prove the uniqueness. We used numeric justifications in Appendices C.3.

Hence, for $-\infty < \lambda^{(12)} < \infty$, $-\infty < \lambda^{(13)} < \infty$ and $w_1 \le -\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon, w_2 \le -\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon$ and $w_3 \le -\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon$,

$$\begin{aligned}\gamma\left(c^*;\lambda^{(12)},\lambda^{(13)},w_1,w_2,w_3,\eta\right) &\le \gamma\left(c^*;\lambda^{(12)},\lambda^{(13)},-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon,-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon,-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon,\eta\right)\\ &\le \gamma\left(c^*;0,0,-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon,-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon,-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon,\eta\right).\end{aligned}$$

Therefore, the type I error in Equation 4.5 can be further simplified to

$$\alpha = \gamma\left(c^*; \lambda^{(12)} = \lambda^{(13)} = 0, w_1 = w_2 = w_3 = -\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon, \eta\right), \tag{4.6}$$

where,

$$\begin{aligned}
&\gamma\left(c^*; \lambda^{(12)} = \lambda^{(13)} = 0, w_1 = w_2 = w_3 = -\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon, \eta\right) \\
=&1 - \sqrt{\frac{3}{4} - \eta^2} \int_{-\infty}^{c^* + \sqrt{\frac{n}{2\sigma_Y^2}}\epsilon - \sqrt{\frac{n}{2\sigma_Y^2}}\epsilon} \phi(z) \int_{-\infty}^{\frac{\eta z}{\sqrt{1-\eta^2}}} \phi(s)\Phi(\frac{\eta z\sqrt{1-\eta^2} - (\frac{1}{2} - \eta^2)s}{\sqrt{\frac{3}{4} - \eta^2}})dsdz \\
&- \sqrt{\frac{3}{4} - \eta^2} \int_{-\infty}^{c^* + \sqrt{\frac{n}{2\sigma_Y^2}}\epsilon - \sqrt{\frac{n}{2\sigma_Y^2}}\epsilon} \phi(z) \int_{-\infty}^{\frac{\eta z}{\sqrt{1-\eta^2}}} \phi(s)\Phi(\frac{\eta z\sqrt{1-\eta^2} - (\frac{1}{2} - \eta^2)s}{\sqrt{\frac{3}{4} - \eta^2}})dsdz \\
&- \sqrt{\frac{3}{4} - \eta^2} \int_{-\infty}^{c^* + \sqrt{\frac{n}{2\sigma_Y^2}}\epsilon - \sqrt{\frac{n}{2\sigma_Y^2}}\epsilon} \phi(z) \int_{-\infty}^{\frac{\eta z}{\sqrt{1-\eta^2}}} \phi(s)\Phi(\frac{\eta z\sqrt{1-\eta^2} - (\frac{1}{2} - \eta^2)s}{\sqrt{\frac{3}{4} - \eta^2}})dsdz \\
=&1 - 3\sqrt{\frac{3}{4} - \eta^2} \int_{-\infty}^{c^*} \phi(z) \int_{-\infty}^{\frac{\eta z}{\sqrt{1-\eta^2}}} \phi(s)\Phi(\frac{\eta z\sqrt{1-\eta^2} - (\frac{1}{2} - \eta^2)s}{\sqrt{\frac{3}{4} - \eta^2}})dsdz.
\end{aligned}$$

Given type I error $\alpha$ and $\eta$, we can determine the critical value $c^*$ by the following equation:

$$\alpha = 1 - 3\sqrt{\frac{3}{4} - \eta^2} \int_{-\infty}^{c^*} \phi(z) \int_{-\infty}^{\frac{\eta z}{\sqrt{1-\eta^2}}} \phi(s)\Phi(\frac{\eta z\sqrt{1-\eta^2} - (\frac{1}{2} - \eta^2)s}{\sqrt{\frac{3}{4} - \eta^2}})dsdz. \tag{4.7}$$

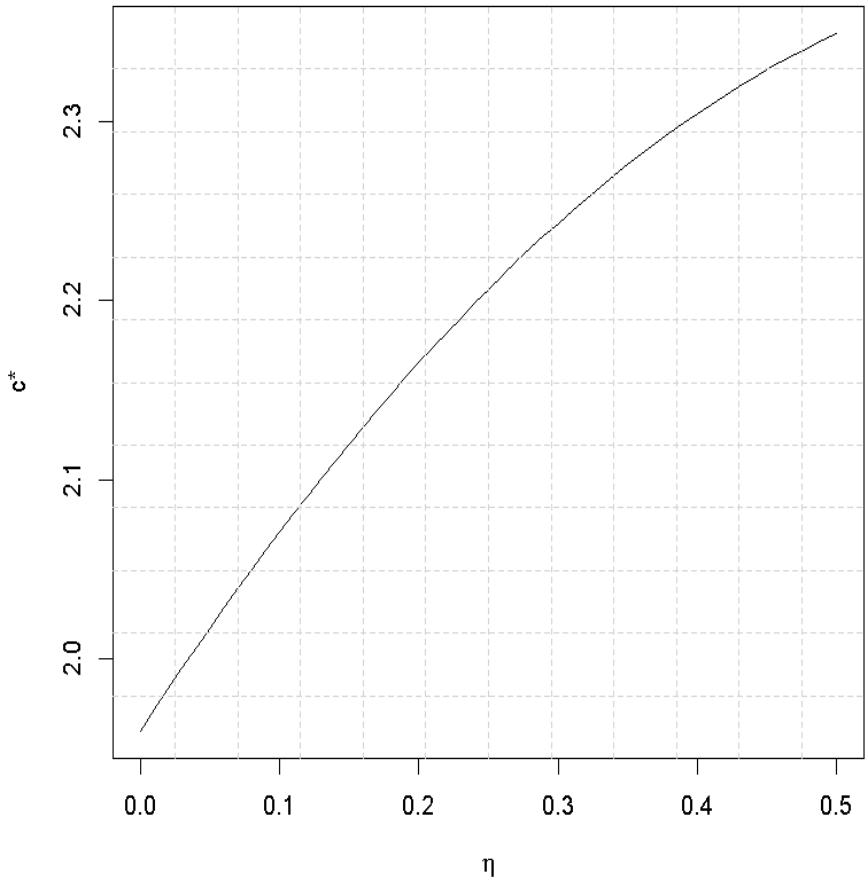
Table 4.1 shows some selective critical values $c^*$ by $\eta$ when $\alpha = 0.025$. Figure 4.1 shows the critical value $c^*$ versus $\eta$ when $\alpha = 0.025$. Since $\eta$ is a function of $\tau$ and $\rho$, for different combinations of $\tau$ and $\rho$, as long as the corresponding $\eta$'s are the same, the critical value will be the same for a given $\alpha$.

#### 4.3.1.2 Superiority Hypothesis

When $\epsilon = 0$, all the properties for the $\gamma(b; \lambda^{(12)}, \lambda^{(13)}, w_1, w_2, w_3)$ function can be applied to the case for superiority hypothesis. Therefore, given the type I error $\alpha$ and $\eta$, the critical value $c$ can be derived by solving:

$$\alpha = 1 - 3\sqrt{\frac{3}{4} - \eta^2} \int_{-\infty}^{c} \phi(z) \int_{-\infty}^{\frac{\eta z}{\sqrt{1-\eta^2}}} \phi(s)\Phi(\frac{\eta z\sqrt{1-\eta^2} - (\frac{1}{2} - \eta^2)s}{\sqrt{\frac{3}{4} - \eta^2}})dsdz. \tag{4.8}$$

Figure 4.1: Critical value $c^*$ versus $\eta$ at $\alpha = 0.025$ when the interim and the final endpoints are different.



Notice that the Equation 4.8 to solve the critical value $c$ for superiority hypothesis and the Equation 4.7 to solve the critical value $c^*$ for non-inferiority are exactly the same. Hence the critical values for controlling the type I error in a two-stage winner design are the same for superiority and for non-inferiority hypotheses. The critical value $c$ or $c^*$ depend on the value of $\eta$, which is a function of information time $\tau$ and correlation $\rho$ between the interim and the final endpoints.

### 4.3.2 Same Endpoints at Interim and Final Analyses

#### 4.3.2.1 Non-inferiority Hypothesis

When same endpoints at interim and final analysis, we have

$$\lambda^{(12)} = \sqrt{\tau}(w_1 - w_2) = 2\eta(w_1 - w_2),$$
$$\lambda^{(13)} = \sqrt{\tau}(w_1 - w_3) = 2\eta(w_1 - w_3).$$

Given $\epsilon$, the type I error of size $\alpha$ test associated with $W^*$ is,

$$\begin{aligned}\alpha &= \sup_{\delta_1 \le -\epsilon, \delta_2 \le -\epsilon, \delta_3 \le -\epsilon} Pr(W^* > c^*; \delta_1, \delta_2, \delta_3, \eta) \\ &= \sup_{w_1 \le -\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon, w_2 \le -\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon, w_3 \le -\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon} Pr(W^* > c^*; w_1, w_2, w_3, \eta). \end{aligned} \tag{4.9}$$

To determine $c^*$, we define a function $\gamma$ as:

$$\begin{aligned}\gamma(b; w_1, w_2, w_3, \eta) &= Pr\left(W^* > b; w_1, w_2, w_3, \eta\right) \\ &= 1 - \sqrt{\frac{3}{4} - \eta^2} \int_{-\infty}^{b - w_1 - \sqrt{\frac{n}{2\sigma_Y^2}}\epsilon} \phi(z) \int_{-\infty}^{\frac{2\eta(w_1 - w_2) + \eta z}{\sqrt{1-\eta^2}}} \phi(s) \\ &\quad \times \Phi\left(\frac{(2\eta(w_1 - w_3) + \eta z)\sqrt{1-\eta^2} - (\frac{1}{2} - \eta^2)s}{\sqrt{\frac{3}{4} - \eta^2}}\right) ds dz \\ &\quad - \sqrt{\frac{3}{4} - \eta^2} \int_{-\infty}^{b - w_2 - \sqrt{\frac{n}{2\sigma_Y^2}}\epsilon} \phi(z) \int_{-\infty}^{\frac{2\eta(w_2 - w_1) + \eta z}{\sqrt{1-\eta^2}}} \phi(s) \\ &\quad \times \Phi\left(\frac{(2\eta(w_2 - w_3) + \eta z)\sqrt{1-\eta^2} - (\frac{1}{2} - \eta^2)s}{\sqrt{\frac{3}{4} - \eta^2}}\right) ds dz\end{aligned}$$

$$- \sqrt{\frac{3}{4} - \eta^2} \int_{-\infty}^{b - w_3 - \sqrt{\frac{n}{2\sigma_Y^2}}\epsilon} \phi(z) \int_{-\infty}^{\frac{2\eta(w_3 - w_1) + \eta z}{\sqrt{1-\eta^2}}} \phi(s)$$
$$\times \Phi(\frac{(2\eta(w_3 - w_2) + \eta z)\sqrt{1-\eta^2} - (\frac{1}{2} - \eta^2)s}{\sqrt{\frac{3}{4} - \eta^2}}) ds dz,$$

where $b$ is some cut-off value. The type I error rate in Equation 4.9 can be rewritten as:

$$\alpha = \sup_{w_1 \leq -\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon, w_2 \leq -\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon, w_3 \leq -\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon} Pr(W^* > c^*; w_1, w_2, w_3, \eta)$$
$$= \sup_{w_1 \leq -\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon, w_2 \leq -\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon, w_3 \leq -\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon} \gamma(c^*; w_1, w_2, w_3, \eta)$$

In order to find the supremum of $\gamma(b; w_1, w_2, w_3)$, we use numerical justification in Appendix C.4 to show the property 4.3.3. We defer the theoretical justification in the future work.

**Property 4.3.3.** *Given $b$, $\gamma(b; w_1, w_2, w_3)$ is monotonically increasing with $w_1, w_2$ and $w_3$ at $\alpha = 0.025, 0.05$.*

Hence under $H_0$, for $w_1 \leq -\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon, w_2 \leq -\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon$ and $w_3 \leq -\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon$, we have

$$\gamma(c^*; w_1, w_2, w_3, \eta) \leq \gamma(c^*; -\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon, -\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon, -\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon, \eta).$$

Therefore, the type I error of size $\alpha$ test can be further simplified to

$$\alpha = \gamma(c^*; w_1 = w_2 = w_3 = -\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon, \eta),$$

where

$$\gamma(c^*; w_1 = w_2 = w_3 = -\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon, \eta)$$
$$= 1 - \sqrt{\frac{3}{4} - \eta^2} \int_{-\infty}^{c^*} \phi(z) \int_{-\infty}^{\frac{\eta z}{\sqrt{1-\eta^2}}} \phi(s) \Phi(\frac{\eta z\sqrt{1-\eta^2} - (\frac{1}{2} - \eta^2)s}{\sqrt{\frac{3}{4} - \eta^2}}) ds dz$$

$$-\sqrt{\frac{3}{4}-\eta^2}\int_{-\infty}^{c^*}\phi(z)\int_{-\infty}^{\frac{\eta z}{\sqrt{1-\eta^2}}}\phi(s)\Phi(\frac{\eta z\sqrt{1-\eta^2}-(\frac{1}{2}-\eta^2)s}{\sqrt{\frac{3}{4}-\eta^2}})dsdz$$
$$-\sqrt{\frac{3}{4}-\eta^2}\int_{-\infty}^{c^*}\phi(z)\int_{-\infty}^{\frac{\eta z}{\sqrt{1-\eta^2}}}\phi(s)\Phi(\frac{\eta z\sqrt{1-\eta^2}-(\frac{1}{2}-\eta^2)s}{\sqrt{\frac{3}{4}-\eta^2}})dsdz.$$

Hence, given type I error $\alpha$ and $\eta$ (or $\tau$), we can determine the critical value $c^*$ by the following equation:

$$\alpha = 1-3\sqrt{\frac{3}{4}-\eta^2}\int_{-\infty}^{c^*}\phi(z)\int_{-\infty}^{\frac{\eta z}{\sqrt{1-\eta^2}}}\phi(s)\Phi(\frac{\eta z\sqrt{1-\eta^2}-(\frac{1}{2}-\eta^2)s}{\sqrt{\frac{3}{4}-\eta^2}})dsdz. \tag{4.10}$$

Table 4.2 shows some critical values $c^*$ with selective information time $\tau$ when $\alpha = 0.025$. Figure 4.2 shows the critical value $c^*$ versus $\tau$ when $\alpha = 0.025$ .

#### 4.3.2.2 Superiority Hypothesis

When $\epsilon = 0$, all the properties for the $\gamma(b; w_1, w_2, w_3)$ function can be applied to the case for superiority hypothesis. Therefore, given the type I error $\alpha$, the critical value $c$ can be derived by solving:

$$\alpha = 1-3\sqrt{\frac{3}{4}-\eta^2}\int_{-\infty}^{c}\phi(z)\int_{-\infty}^{\frac{\eta z}{\sqrt{1-\eta^2}}}\phi(s)\Phi(\frac{\eta z\sqrt{1-\eta^2}-(\frac{1}{2}-\eta^2)s}{\sqrt{\frac{3}{4}-\eta^2}})dsdz. \tag{4.11}$$

Notice that the Equation 4.11 to solve the $c$ for superiority hypothesis and the Equation 4.10 to solve the $c^*$ for non-inferiority are exactly the same. Based on the Equations (4.7), (4.8), (4.10) and (4.11), we have the following findings: 1) Critical value $c$ for superiority trial is the same as the critical value $c^*$ for non-inferiority trial; 2) Critical value will be the same regardless when surrogate endpoint is used or not.

When surrogate endpoint is <u>not</u> used, if a given $\tau$ resulted in the same $\eta$ value as when surrogate endpoint is used, the resulting critical value will be the same. For example, when surrogate endpoint is not used, at $\tau = 0.25$, the corresponding $\eta = 0.25$ with critical value 2.206 (see Table 4.2), which is the same as the critical value when surrogate endpoint is used at $\eta = 0.25$ (see Table 4.1).

Table 4.1: Critical value $c^*$ of two-stage winner designs with superiority or non-inferiority hypothesis with three experimental treatment and one active control arms when $\alpha = 0.025$, and when different endpoints at the interim and the final analyses.

| | $(\tau, \rho)$ | | | | |
|---|---|---|---|---|---|
| | $(0.25, 0)$ | $(0.33, 00)$ | $(0.5, 0)$ | $(0.75, 0)$ | $(1, 0)$ |
| $\eta$ | 0 | 0 | 0 | 0 | 0 |
| $c^*$ | 1.96 | 1.96 | 1.96 | 1.96 | 1.96 |
| | $(\tau, \rho)$ | | | | |
| | $(0.25, 0.2)$ | $(0.33, 0.2)$ | $(0.5, 0.2)$ | $(0.75, 0.2)$ | $(1, 0.2)$ |
| $\eta$ | 0.050 | 0.057 | 0.071 | 0.087 | 0.100 |
| $c^*$ | 2.018 | 2.026 | 2.040 | 2.057 | 2.071 |
| | $(\tau, \rho)$ | | | | |
| | $(0.25, 0.5)$ | $(0.33, 0.5)$ | $(0.5, 0.5)$ | $(0.75, 0.5)$ | $(1, 0.5)$ |
| $\eta$ | 0.125 | 0.144 | 0.177 | 0.217 | 0.250 |
| $c^*$ | 2.096 | 2.114 | 2.145 | 2.179 | 2.206 |
| | $(\tau, \rho)$ | | | | |
| | $(0.25, 0.8)$ | $(0.33, 0.8)$ | $(0.5, 0.8)$ | $(0.75, 0.8)$ | $(1, 0.8)$ |
| $\eta$ | 0.200 | 0.230 | 0.283 | 0.346 | 0.400 |
| $c^*$ | 2.165 | 2.190 | 2.231 | 2.274 | 2.304 |

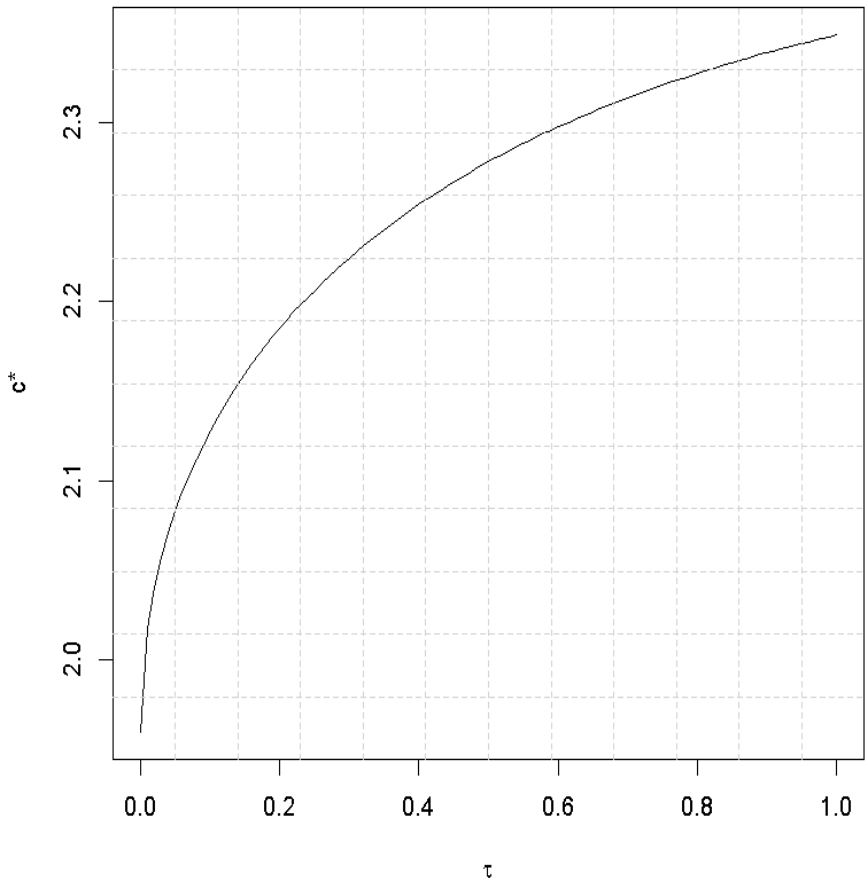
With given $\tau$ and $\rho$, $\eta = \frac{\sqrt{\tau}}{2}\rho$ can be determined.

Table 4.2: Critical value $c^*$ of the two-stage winner design for superiority or non-inferiority hypothesis with three experimental treatment and one active control arms when $\alpha = 0.025$ and when same endpoints at the interim and the final analyses.

| | $\tau$ | | | | |
|---|---|---|---|---|---|
| | 0.25 | 0.33 | 0.5 | 0.75 | 1 |
| $\eta$ | 0.250 | 0.287 | 0.354 | 0.433 | 0.500 |
| $c^*$ | 2.206 | 2.234 | 2.278 | 2.321 | 2.349 |

With given information time $\tau$, the $\eta$ can be determined by $\eta = \frac{\sqrt{\tau}}{2}$.

Figure 4.2: Critical value $c^*$ versus $\tau$ when $\alpha = 0.025$ and when same endpoints at the interim and the final analyses.



## 4.4 Power

In this section, we discuss the power of the test.

### 4.4.1 Different Endpoints at Interim and Final Analyses

When different endpoints are used at the interim and the final analyses, we define the power $1-\beta$ as follow:

$$
\begin{aligned}
1-\beta =& \Pr\left(W^* > c^*; H_1\right) \\
=& \gamma(c^*; \lambda^{(12)}, \lambda^{(13)}, w_1, w_2, w_3, \eta) \\
=& 1 - \sqrt{\frac{3}{4}-\eta^2} \int_{-\infty}^{c^*-w_1-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon} \phi(z) \int_{-\infty}^{\frac{\lambda^{(12)}+\eta z}{\sqrt{1-\eta^2}}} \phi(s) \\
& \times \Phi\left(\frac{\left(\lambda^{(13)}+\eta z\right)\sqrt{1-\eta^2} - (\frac{1}{2}-\eta^2)s}{\sqrt{\frac{3}{4}-\eta^2}}\right) ds dz \\
& - \sqrt{\frac{3}{4}-\eta^2} \int_{-\infty}^{c^*-w_2-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon} \phi(z) \int_{-\infty}^{\frac{-\lambda^{(12)}+\eta z}{\sqrt{1-\eta^2}}} \phi(s) \\
& \times \Phi\left(\frac{\left(\lambda^{(13)}-\lambda^{(12)}+\eta z\right)\sqrt{1-\eta^2} - (\frac{1}{2}-\eta^2)s}{\sqrt{\frac{3}{4}-\eta^2}}\right) ds dz \\
& - \sqrt{\frac{3}{4}-\eta^2} \int_{-\infty}^{c^*-w_3-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon} \phi(z) \int_{-\infty}^{\frac{-\lambda^{(13)}+\eta z}{\sqrt{1-\eta^2}}} \phi(s) \\
& \times \Phi\left(\frac{\left(\lambda^{(12)}-\lambda^{(13)}+\eta z\right)\sqrt{1-\eta^2} - (\frac{1}{2}-\eta^2)s}{\sqrt{\frac{3}{4}-\eta^2}}\right) ds dz. \qquad (4.12)
\end{aligned}
$$

When $\epsilon = 0$, Equation 4.12 becomes the power function for superiority hypothesis. Given type I error $\alpha$ and $\eta$, critical value $c^*$ can be obtained from Equation 4.7. In the following paragraphs, we discuss the power function under three scenarios.

- Case I: $\delta_1 = \delta_2 = \delta_3 = \delta$.

  When $\delta_1 = \delta_2 = \delta_3 = \delta$, the power $1-\beta$ can be calculated as:

$$
\begin{aligned}
1-\beta =& \Pr\left(W^* > c^*; H_1\right) \\
=& \gamma(c^*; \lambda^{(12)}, \lambda^{(13)}, w_1 = w_2 = w_3 = \sqrt{\frac{n}{2\sigma_Y^2}}\delta, \eta)
\end{aligned}
$$

$$
\begin{aligned}
=&1-\sqrt{\frac{3}{4}-\eta^2}\int_{-\infty}^{c^*-\sqrt{\frac{n}{2\sigma_Y^2}}(\delta+\epsilon)}\phi(z)\int_{-\infty}^{\frac{\lambda^{(12)}+\eta z}{\sqrt{1-\eta^2}}}\phi(s)\\
&\times\Phi(\frac{\left(\lambda^{(13)}+\eta z\right)\sqrt{1-\eta^2}-(\frac{1}{2}-\eta^2)s}{\sqrt{\frac{3}{4}-\eta^2}})dsdz\\
&-\sqrt{\frac{3}{4}-\eta^2}\int_{-\infty}^{c^*-\sqrt{\frac{n}{2\sigma_Y^2}}(\delta+\epsilon)}\phi(z)\int_{-\infty}^{\frac{-\lambda^{(12)}+\eta z}{\sqrt{1-\eta^2}}}\phi(s)\\
&\times\Phi(\frac{\left(\lambda^{(13)}-\lambda^{(12)}+\eta z\right)\sqrt{1-\eta^2}-(\frac{1}{2}-\eta^2)s}{\sqrt{\frac{3}{4}-\eta^2}})dsdz\\
&-\sqrt{\frac{3}{4}-\eta^2}\int_{-\infty}^{c^*-\sqrt{\frac{n}{2\sigma_Y^2}}(\delta+\epsilon)}\phi(z)\int_{-\infty}^{\frac{-\lambda^{(13)}+\eta z}{\sqrt{1-\eta^2}}}\phi(s)\\
&\times\Phi(\frac{\left(\lambda^{(12)}-\lambda^{(13)}+\eta z\right)\sqrt{1-\eta^2}-(\frac{1}{2}-\eta^2)s}{\sqrt{\frac{3}{4}-\eta^2}})dsdz \qquad (4.13)
\end{aligned}
$$

From Equation 4.13, the power is a function of $c^*, \delta, \epsilon, \sigma_Y, \lambda^{(12)}, \lambda^{(13)}$, $n$ and $\eta$ when different endpoints at the interim and the final analyses. For different combinations of $\delta$ and $\epsilon$, as long as the sum of these two parameters are the same, so is the power. For instance, when $\delta = 0.1, \epsilon = 0.1$ the power will be the same as when $\delta = 0, \epsilon = 0.2$.

To study the effect of $\tau, n$ and $\epsilon$ on power, we show the power curves for $\delta_1 = \delta_2 = \delta_3 = \delta$ when $\nu_1^X - \nu_2^X = 0, \sigma_X = \sigma_Y = 1, \alpha = 0.025, \nu_1^X - \nu_3^X = 0$, and $\rho = 0.5$ in Figure 4.3. When $\tau$ increases, there is no changes on power in the figures. When $\epsilon$ or $n$ increases, the power increase.

To study the effect of $\rho, n$ and $\epsilon$ on power, we show the power curves for $\delta_1 = \delta_2 = \delta_3 = \delta$ when $\nu_1^X - \nu_2^X = 0, \nu_1^X - \nu_3^X = 0, \alpha = 0.025, \sigma_X = \sigma_Y = 1$ and $\tau = 0.5$ in Figure 4.4. As seen in the figures, $\rho$ have minimum effect on the power. However, it is obvious that when $\epsilon$ or $n$ increases, the power curves shift to the left suggesting power increase.

To study the effect of $\nu_1^X - \nu_2^X$ and $\nu_1^X - \nu_3^X$ on power, we show the power curves for $\delta_1 = \delta_2 = \delta_3 = \delta$ when $\nu_1^X - \nu_2^X = 0.1, \nu_1^X - \nu_3^X = 0.3, \alpha = 0.025, \sigma_X = \sigma_Y = 1, \rho = 0.2, 0.5, 0.8, \epsilon = 0, 0.1, 0.2$ and $\tau = 0.5$ in Figure 4.5. The only difference between Figure 4.5 and 4.4 is the $\nu_1^X - \nu_2^X$ and $\nu_1^X - \nu_3^X$ at the interim look. Comparing the Figure 4.5 and 4.4, when $\nu_1^X - \nu_2^X$ increases $\nu_1^X - \nu_3^X$, there is difference in these

Figure 4.3: Power curves for $\delta_1 = \delta_2 = \delta_3 = \delta$ when $\nu_1^X - \nu_2^X = 0, \nu_1^X - \nu_3^X = 0, \alpha = 0.025, \sigma_X = \sigma_Y = 1$ and $\rho = 0.5$ by $\tau, n$ and $\epsilon$ when different endpoints at interim and final analyses.

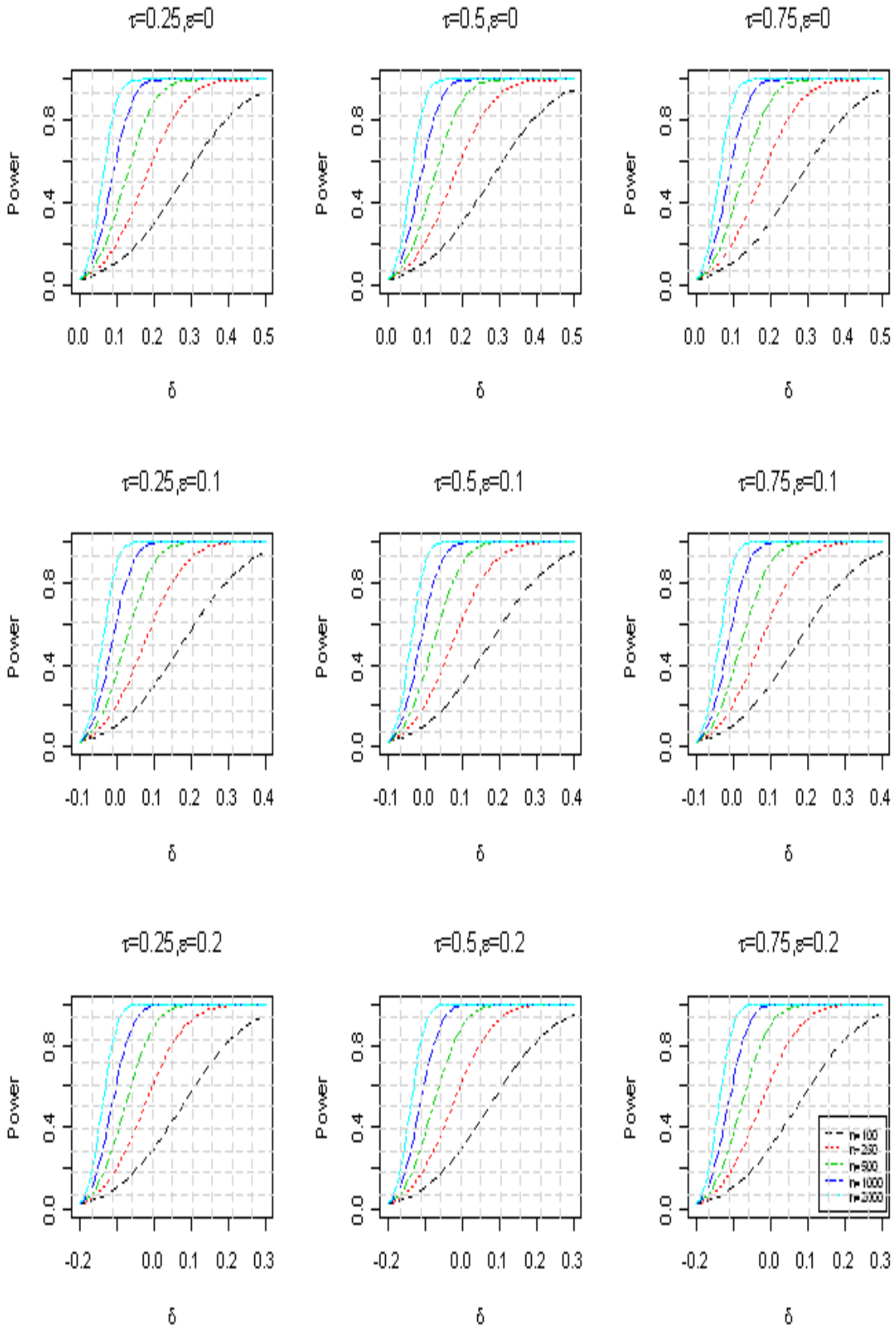


Figure 4.4: Power curves for $\delta_1 = \delta_2 = \delta_3 = \delta$ when $\nu_1^X - \nu_2^X = 0, \nu_1^X - \nu_3^X = 0, \alpha = 0.025, \sigma_X = \sigma_Y = 1$ and $\tau = 0.5$ by $\rho, n$ and $\epsilon$ when different endpoints at interim and final analyses.

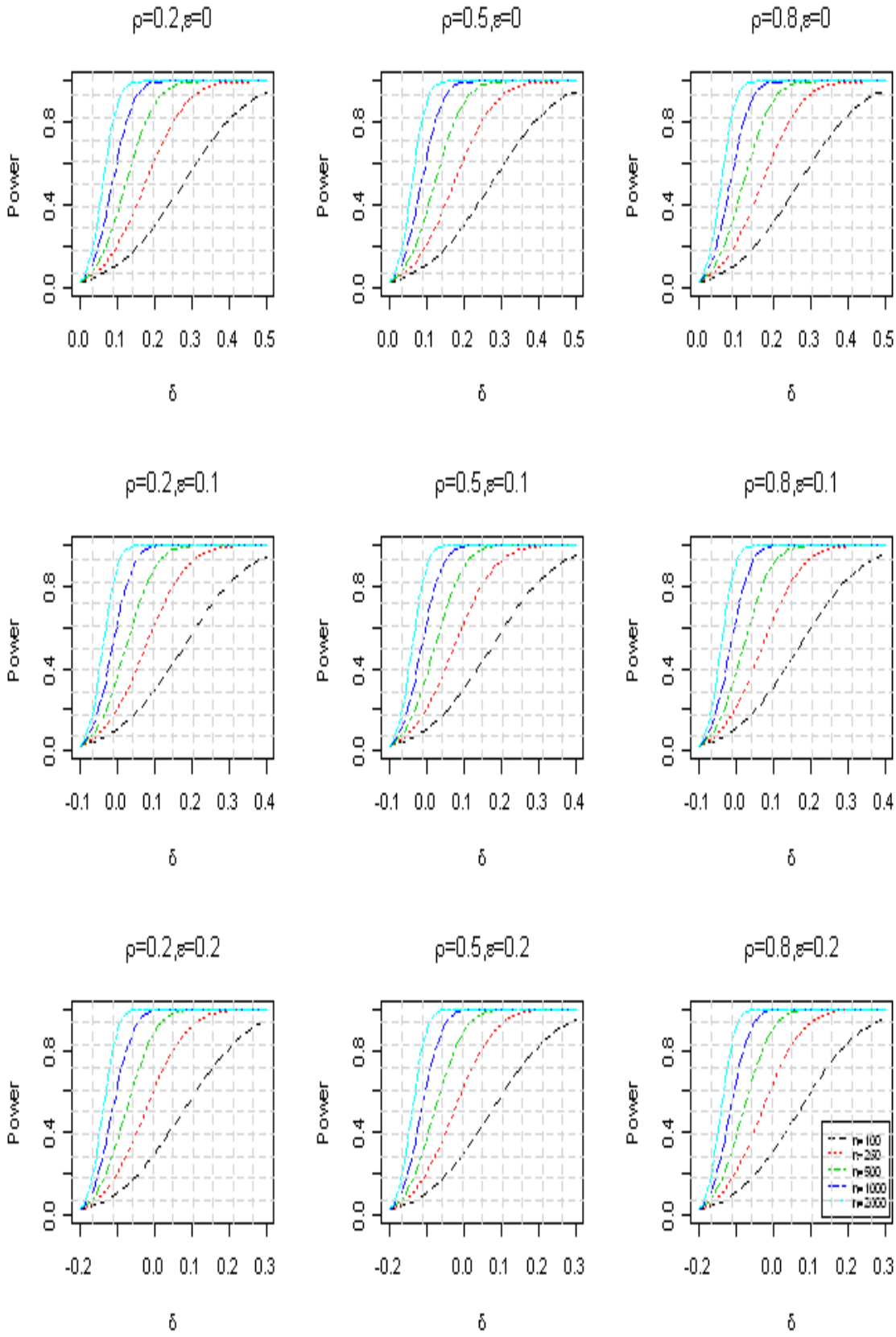


Table 4.3: Probability of selecting each experimental treatment, by information time $\tau$, $\nu_1^X - \nu_2^X$ and $\nu_1^X - \nu_3^X$.

| | $\nu_1^X - \nu_2^X = 0$, $\nu_1^X - \nu_3^X = 0$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | $\tau = 0.25$ | | | $\tau = 0.5$ | | | $\tau = 0.75$ | | |
| n | 100 | 250 | 500 | 100 | 250 | 500 | 100 | 250 | 500 |
| $p_1$ | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 |
| $p_2$ | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 |
| $p_3$ | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 |
| | $\nu_1^X - \nu_2^X = 0.1$, $\nu_1^X - \nu_3^X = 0.3$ | | | | | | | | |
| | $\tau = 0.25$ | | | $\tau = 0.5$ | | | $\tau = 0.75$ | | |
| n | 100 | 250 | 500 | 100 | 250 | 500 | 100 | 250 | 500 |
| $p_1$ | 0.59 | 0.70 | 0.78 | 0.67 | 0.78 | 0.87 | 0.72 | 0.83 | 0.91 |
| $p_2$ | 0.33 | 0.28 | 0.21 | 0.30 | 0.21 | 0.13 | 0.26 | 0.17 | 0.09 |
| $p_3$ | 0.08 | 0.02 | 0.004 | 0.03 | 0.004 | $< 0.001$ | 0.01 | $< 0.001$ | $< 0.001$ |
| | $\nu_1^X - \nu_2^X = 0$, $\nu_1^X - \nu_3^X = 0.3$ | | | | | | | | |
| | $\tau = 0.25$ | | | $\tau = 0.5$ | | | $\tau = 0.75$ | | |
| n | 100 | 250 | 500 | 100 | 250 | 500 | 100 | 250 | 500 |
| $p_1$ | 0.47 | 0.49 | 0.50 | 0.49 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 |
| $p_2$ | 0.47 | 0.49 | 0.50 | 0.49 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 |
| $p_3$ | 0.05 | 0.01 | 0.001 | 0.02 | 0.001 | $< 0.001$ | 0.007 | 0.0001 | 0.0001 |

power curves. This is because when $\delta_1 = \delta_2 = \delta_3 = \delta$, and the endpoints at the interim and the final are different, the power does not depend on the treatment effect of the surrogate endpoint at interim analysis.

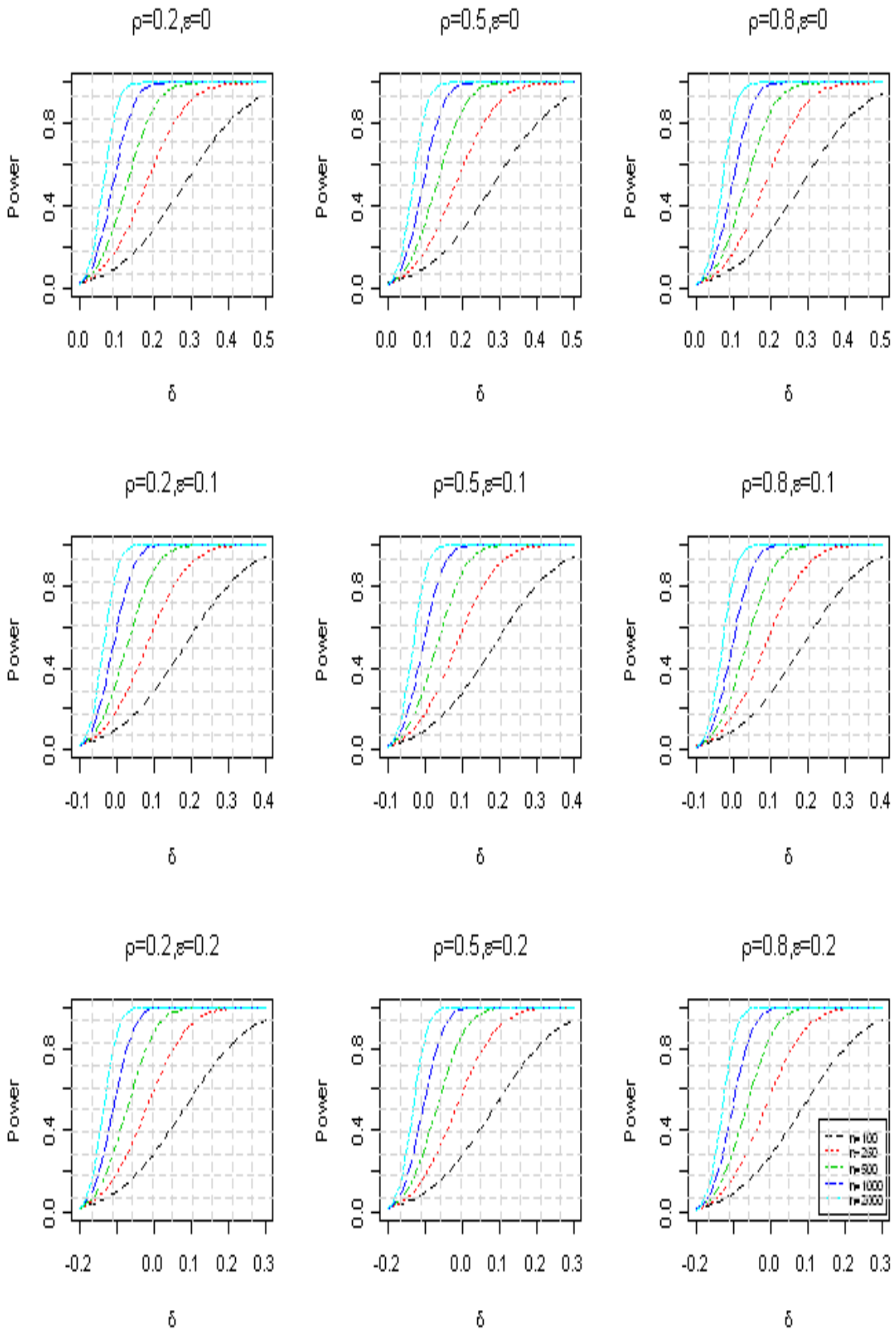
Figures 4.3, 4.4 and 4.5 show that when $\delta_1 = \delta_2 = \delta_3 = \delta$, the power does not depend on the $\tau$, $\rho$, $\nu_1^X - \nu_2^X = 0.1$ or $\nu_1^X - \nu_3^X = 0.3$. This is clear since there is no differences between the experimental treatment arms and the active control at the final analysis, regardless when or which treatment we select, it have no impact on the power.

- Case II: $\delta_1 = \delta_2 = \delta \neq \delta_3$.

  When $\delta_1 = \delta_2 = \delta \neq \delta_3$, the power can be calculated as:

$$
\begin{aligned}
1 - \beta =& \Pr\left(W^* > c^*; H_1\right) \\
=& \gamma(c^*; \lambda^{(12)}, \lambda^{(13)}, w_1 = w_2 = \sqrt{\frac{n}{2\sigma_Y^2}}\delta, w_3 = \sqrt{\frac{n}{2\sigma_Y^2}}\delta_3, \eta) \\
=& 1 - \sqrt{\frac{3}{4} - \eta^2} \int_{-\infty}^{c^* - \sqrt{\frac{n}{2\sigma_Y^2}}(\delta+\epsilon)} \phi(z) \int_{-\infty}^{\frac{\lambda^{(12)} + \eta z}{\sqrt{1-\eta^2}}} \phi(s)
\end{aligned}
$$

Figure 4.5: Power curve for $\delta_1 = \delta_2 = \delta_3 = \delta$ when $\nu_1^X - \nu_2^X = 0.1, \nu_1^X - \nu_3^X = 0.3, \alpha = 0.025, \sigma_X = \sigma_Y = 1$ and $\tau = 0.5$ by $\rho, n$ and $\epsilon$ when different endpoints at interim and final analyses.



$$\times \Phi\left(\frac{\left(\lambda^{(13)}+\eta z\right)\sqrt{1-\eta^2}-(\frac{1}{2}-\eta^2)s}{\sqrt{\frac{3}{4}-\eta^2}}\right)dsdz$$

$$-\sqrt{\frac{3}{4}-\eta^2}\int_{-\infty}^{c^*-\sqrt{\frac{n}{2\sigma_Y^2}}(\delta+\epsilon)}\phi(z)\int_{-\infty}^{\frac{-\lambda^{(12)}+\eta z}{\sqrt{1-\eta^2}}}\phi(s)$$

$$\times \Phi\left(\frac{\left(\lambda^{(13)}-\lambda^{(12)}+\eta z\right)\sqrt{1-\eta^2}-(\frac{1}{2}-\eta^2)s}{\sqrt{\frac{3}{4}-\eta^2}}\right)dsdz$$

$$-\sqrt{\frac{3}{4}-\eta^2}\int_{-\infty}^{c^*-\sqrt{\frac{n}{2\sigma_Y^2}}(\delta_3+\epsilon)}\phi(z)\int_{-\infty}^{\frac{-\lambda^{(13)}+\eta z}{\sqrt{1-\eta^2}}}\phi(s)$$

$$\times \Phi\left(\frac{\left(\lambda^{(12)}-\lambda^{(13)}+\eta z\right)\sqrt{1-\eta^2}-(\frac{1}{2}-\eta^2)s}{\sqrt{\frac{3}{4}-\eta^2}}\right)dsdz. \tag{4.14}$$

From Equation 4.14, the power is a function of $c^*, \delta_1 = \delta_2 = \delta, \delta_3, \epsilon, \sigma_Y, \lambda^{(12)}, \lambda^{(13)}, n$ and $\eta$.

Figure 4.6 shows the effect of $\nu_1^X - \nu_2^X = \nu_{12}$ and $\nu_1^X - \nu_3^X = \nu_{13}$ on the power when $\delta_1 = \delta_2 = \delta \neq \delta_3$, $\tau = 1, \rho = 0.5, n = 250, \sigma_X = \sigma_Y = 1, \epsilon = 0$ and $\alpha = 0.025$. If the treatment effects at the interim is in the same direction as the final endpoints, the power will be bigger as compare to when it's in the opposite direction. For example, when $\delta_1 = \delta_2 = \delta > \delta_3$, the power will be bigger if $\nu_{12} < \nu_{13}$ than if $\nu_{12} > \nu_{13}$.

Table 4.3 provides the probability of selecting each arm by $\tau, n, \nu_1^X - \nu_2^X$ and $\nu_1^X - \nu_3^X$. When $\nu_1^X - \nu_2^X > 0$ and $\nu_1^X - \nu_3^X > 0$, regardless the size of $n$ or $\tau$, the probability of selecting treatment 1 or 2 is way higher than selecting treatment 3. Therefore, as seen in the figures, the power depend mainly on the effect of $\delta_1 = \delta_2 = \delta$, but not $\delta_3$.

Figure 4.7 shows the effects of $\tau$ and $\rho$ on power when $\delta_1 = \delta_2 = \delta \neq \delta_3, n = 250, \epsilon = 0, \nu_1^X - \nu_2^X = 0, \nu_1^X - \nu_3^X = 0.3, \sigma_X = \sigma_Y = 1$ and $\alpha = 0.025$. When $\tau$ changes, the power did not change much. When $\nu_1^X - \nu_2^X = 0, \nu_1^X - \nu_3^X = 0.3$, and $n = 250$, we know the probability of selecting treatment 1 or 2 are almost 0.5, regardless when we perform the interim selection (see Table 4.3). Therefore, when $\tau$ increase, we do not see much increase on the power. Furthermore, as $\rho$ changes, the power surface did not change much suggesting it have minimum impact on the power.

Figure 4.6: Contour of power surface for $\delta_1 = \delta_2 = \delta \neq \delta_3$, when $\alpha = 0.025, \tau = 1, \rho = 0.5, \sigma_X = \sigma_Y = 1, \epsilon = 0$ and $n = 250$ by $\nu_{12} = \nu_1^X - \nu_2^X$ and $\nu_{13} = \nu_1^X - \nu_3^X$ when different endpoints at interim and final analyses.

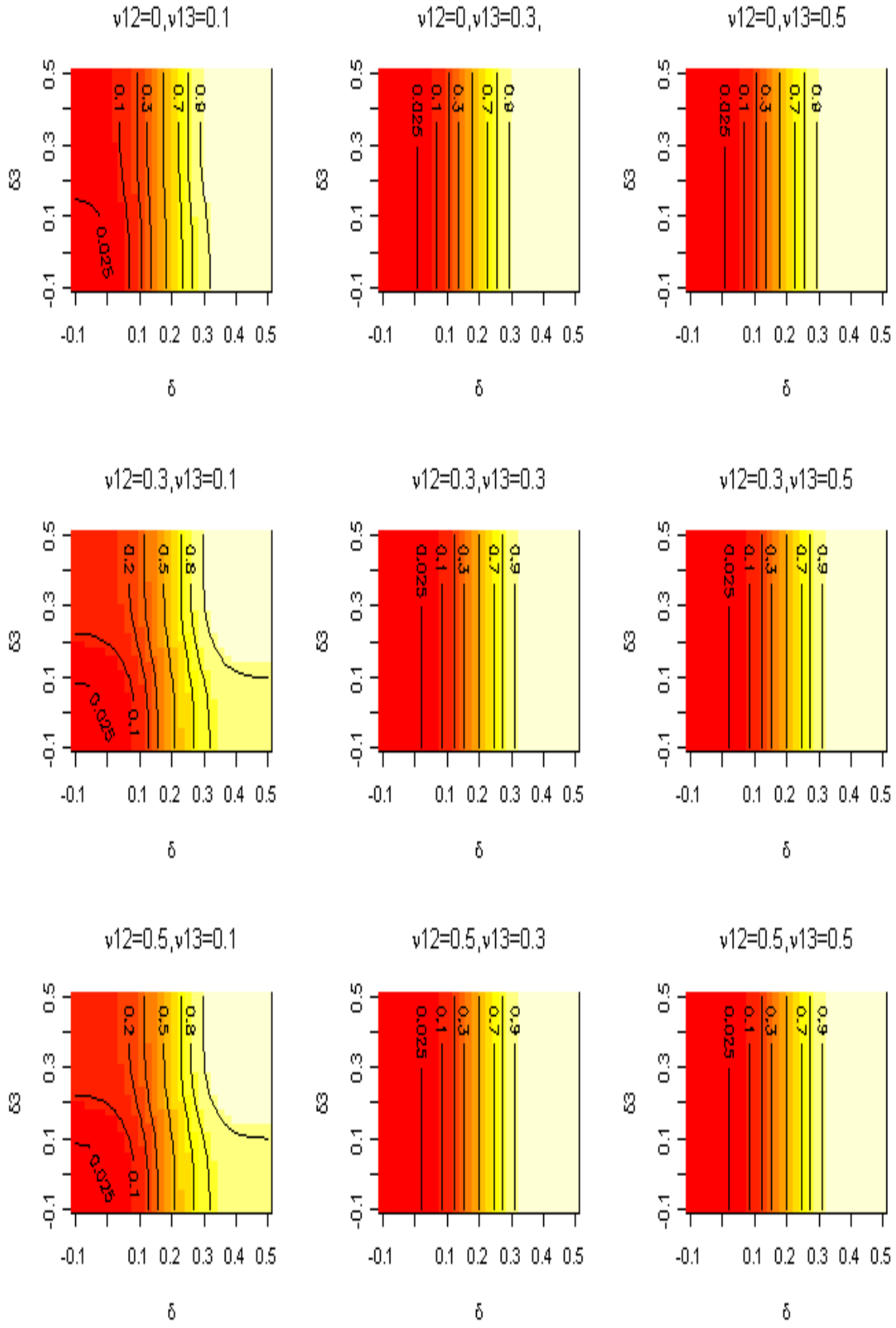


Figure 4.7: Contour of power surface for $\delta_1 = \delta_2 = \delta \neq \delta_3$ when $n = 250$, $\alpha = 0.025, \epsilon = 0, \nu_1^X - \nu_2^X = 0, \sigma_X = \sigma_Y = 1$ and $\nu_1^X - \nu_3^X = 0.3$ by $\tau$ and $\rho$ when different endpoints at interim and final analyses.

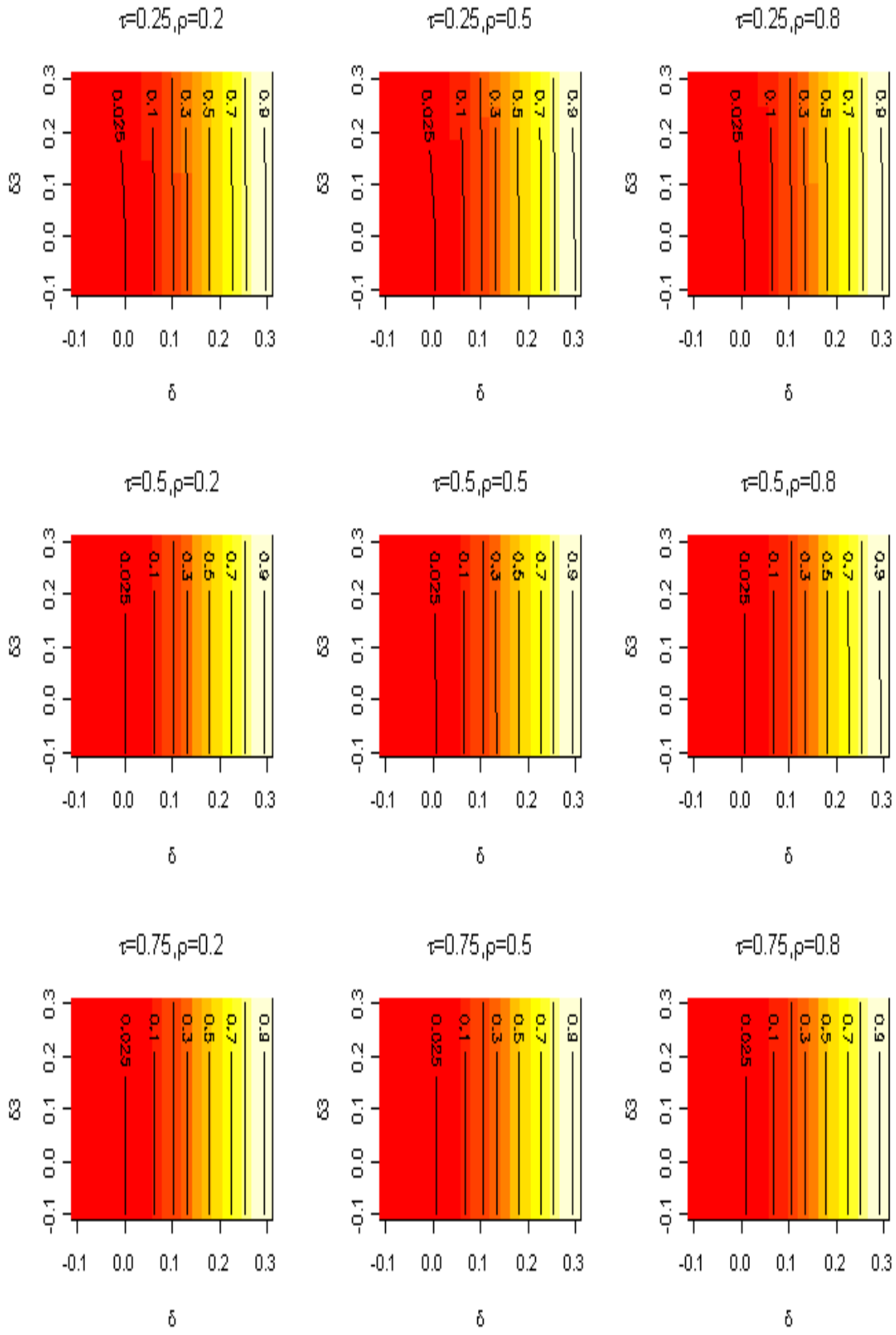


Figure 4.8 shows the effects of $\tau$ and $\epsilon$ on power for $\delta_1 = \delta_2 \neq \delta_3$ when $\nu_1^X - \nu_2^X = 0$, $\nu_1^X - \nu_3^X = 0.3, \rho = 0.8, \alpha = 0.025, \sigma_X = \sigma_Y = 1$ and $n = 250$. When $\tau$ changes, we can hardly see any changes in the power. This is because when $\nu_1^X - \nu_2^X = 0$, $\nu_1^X - \nu_3^X = 0.3$, there is an equal probability of selecting the treatment 1 or 2, and that probability is almost 0.5(see Table 4.3). Therefore, regardless when we perform the interim selection, the probability of selecting the treatment 1 and 2 will not change much. As a result, the power surface does not depend on $\delta_3$ very much. Furthermore, as either $\epsilon$ or $\delta$ increases, the power increases.

- Case III: $\delta_1 \neq \delta_2 \neq \delta_3$.

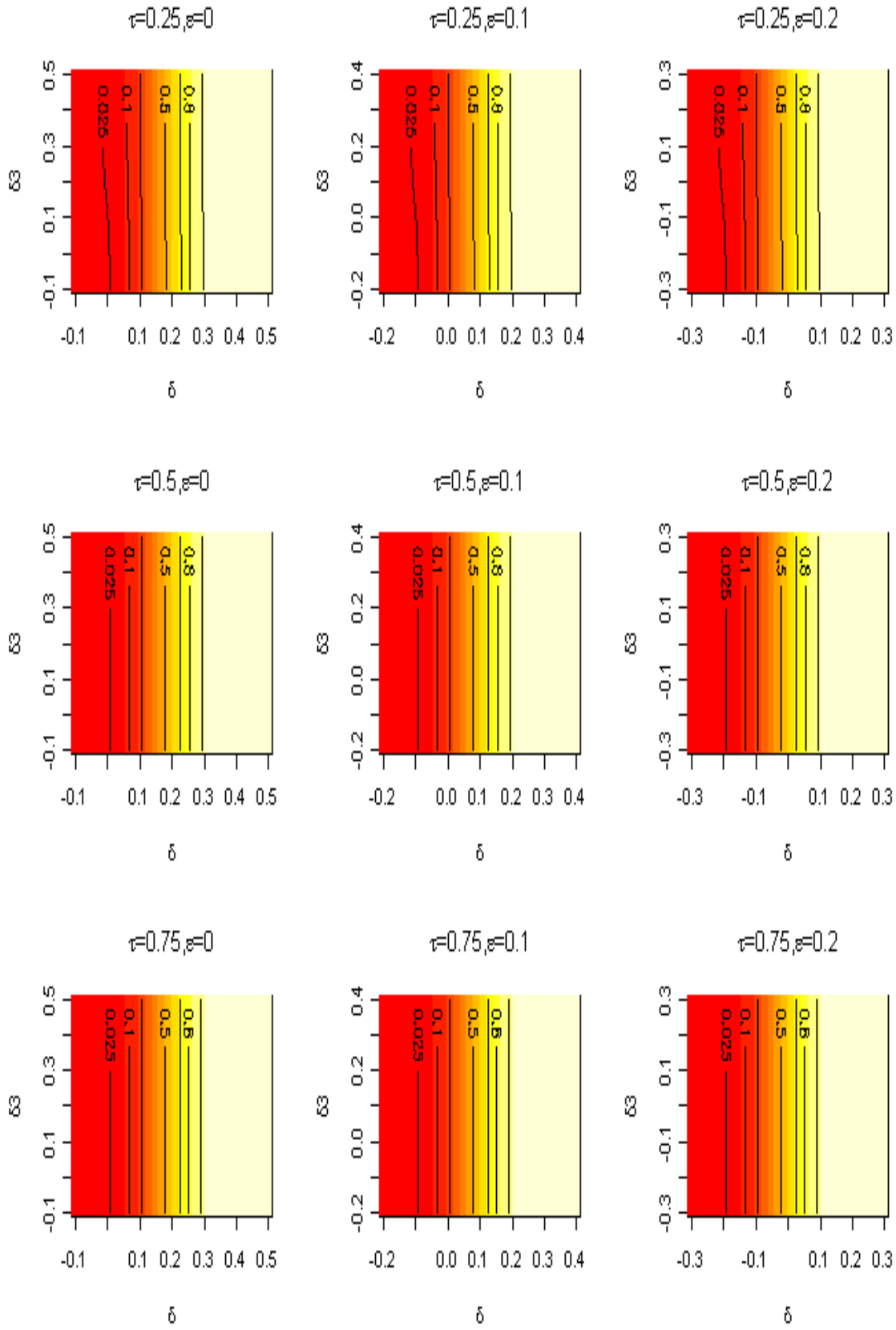
  When $\delta_1 \neq \delta_2 \neq \delta_3$, the power can be calculated by

$$
\begin{aligned}
1-\beta =& \Pr\left(W^* > c^*; H_1\right) \\
=& \gamma(c^*; \lambda^{(12)}, \lambda^{(13)}, w_1 = \sqrt{\frac{n}{2\sigma_Y^2}}\delta_1, w_2 = \sqrt{\frac{n}{2\sigma_Y^2}}\delta_2, w_3 = \sqrt{\frac{n}{2\sigma_Y^2}}\delta_3, \eta) \\
=& 1 - \sqrt{\frac{3}{4}-\eta^2}\int_{-\infty}^{c^* - \sqrt{\frac{n}{2\sigma_Y^2}}(\delta_1+\epsilon)} \phi(z) \int_{-\infty}^{\frac{\lambda^{(12)}+\eta z}{\sqrt{1-\eta^2}}} \phi(s) \\
& \times \Phi\left(\frac{\left(\lambda^{(13)} + \eta z\right)\sqrt{1-\eta^2} - (\frac{1}{2}-\eta^2)s}{\sqrt{\frac{3}{4}-\eta^2}}\right) ds dz \\
& - \sqrt{\frac{3}{4}-\eta^2}\int_{-\infty}^{c^* - \sqrt{\frac{n}{2\sigma_Y^2}}(\delta_2+\epsilon)} \phi(z) \int_{-\infty}^{\frac{-\lambda^{(12)}+\eta z}{\sqrt{1-\eta^2}}} \phi(s) \\
& \times \Phi\left(\frac{\left(\lambda^{(13)} - \lambda^{(12)} + \eta z\right)\sqrt{1-\eta^2} - (\frac{1}{2}-\eta^2)s}{\sqrt{\frac{3}{4}-\eta^2}}\right) ds dz \\
& - \sqrt{\frac{3}{4}-\eta^2}\int_{-\infty}^{c^* - \sqrt{\frac{n}{2\sigma_Y^2}}(\delta_3+\epsilon)} \phi(z) \int_{-\infty}^{\frac{-\lambda^{(13)}+\eta z}{\sqrt{1-\eta^2}}} \phi(s) \\
& \times \Phi\left(\frac{\left(\lambda^{(12)} - \lambda^{(13)} + \eta z\right)\sqrt{1-\eta^2} - (\frac{1}{2}-\eta^2)s}{\sqrt{\frac{3}{4}-\eta^2}}\right) ds dz. \qquad (4.15)
\end{aligned}
$$

From Equation 4.15, the power is a function of $c^*, \delta_1, \delta_2, \delta_3, \sigma_Y, \epsilon, \lambda^{(12)}, \lambda^{(13)}$, $n$ and $\eta$.

Figure 4.9 shows the effects of $\delta_3$ and $n$ on the contour plot of power surface for $\delta_1 \neq \delta_2 \neq \delta_3$ when $\nu_1^X - \nu_2^X = 0.1, \nu_1^X - \nu_3^X = 0.3, \eta = 0.25, \epsilon = 0, \sigma_X = \sigma_Y = 1$ and $\alpha = 0.025$. As $\delta_3$ increases, there is no visual changes on the power surfaces. This

Figure 4.8: Contour of power surface for $\delta_1 = \delta_2 \neq \delta_3$ when $\nu_1^X - \nu_2^X = 0, \nu_1^X - \nu_3^X = 0.3, \alpha = 0.025, \rho = 0.8, \sigma_X = \sigma_Y = 1$ and $n = 250$ by $\tau$ and $\epsilon$ when different endpoints at interim and final analyses.



is because of the parameters that we choose here for the figures. In the figures, we choose $\nu_1^X - \nu_2^X = 0.1$, $\nu_1^X - \nu_3^X = 0.3$ $\tau = 1$ and $\rho = 0.5$, there is slim chance to pick treatment arm 3 (see Table 4.3) at the interim look, therefore, $\delta_3$ did not have much impact on the power. On the other hand, as $n$ increases, power increases.

Figure 4.10 shows the effects of $\tau$ and $\rho$ on the power for $\delta_1 \neq \delta_2 \neq \delta_3$ when $\nu_1^X - \nu_2^X = 0.1$, $\nu_1^X - \nu_3^X = 0.3$, $n = 250, \delta_3 = 0.1, \sigma_X = \sigma_Y = 1, \epsilon = 0$ and $\alpha = 0.025$. As $\tau$ increases, so is the probability of selecting the right winner, therefore, the power increases. On the other hand, as $\rho$ changes, we can hardly see any changes on the power.

Figure 4.11 shows the effects of $\epsilon$ and $n$ on the power surface for $\delta_1 \neq \delta_2 \neq \delta_3$ when $\nu_1^X - \nu_2^X = 0.1, \nu_1^X - \nu_3^X = 0.3, \alpha = 0.025, \tau = 1, \rho = 0.5 (i.e. \eta = 0.25), \sigma_X = \sigma_Y = 1$ and $\delta_3 = 0.1$. As the sample size $n$ increases, the power increases. When $\epsilon$ increases, the power surface shifts to left suggesting power increases.

Figure 4.12 shows the effects of $\tau$ and $\epsilon$ for $\delta_1 \neq \delta_2 \neq \delta_3$ when $\nu_1^X - \nu_2^X = 0.1, \nu_1^X - \nu_3^X = 0.3, \delta_3 = 0.1, n = 250, \alpha = 0.025, \sigma_X = \sigma_Y = 1$ and $\rho = 0.8$. From the figures, when the we perform interim selection at a later time during the trial (i.e. increases $\tau$), the power increases. As $\epsilon$ increases, the power surface shift to left suggesting increases on power.

### 4.4.2 Same Endpoints at Interim and Final Analyses

When same endpoints are used at the interim and final analysis, we can calculate the power $1 - \beta$ as follows:

$$
\begin{aligned}
1 - \beta =& Pr\left(W^* > c^*; H_1\right) \\
=& \gamma(c^*; w_1, w_2, w_3, \eta) \\
=& 1 - \sqrt{\frac{3}{4} - \eta^2} \int_{-\infty}^{c^* - w_1 - \sqrt{\frac{n}{2\sigma_Y^2}}\epsilon} \phi(z) \int_{-\infty}^{\frac{\sqrt{\tau}(w_1 - w_2) + \eta z}{\sqrt{1-\eta^2}}} \phi(s) \\
& \times \Phi\left(\frac{\left(\sqrt{\tau}(w_1 - w_3) + \eta z\right)\sqrt{1-\eta^2} - (\frac{1}{2} - \eta^2)s}{\sqrt{\frac{3}{4} - \eta^2}}\right) ds dz
\end{aligned}
$$

Figure 4.9: Contour of power surface for $\delta_1 \neq \delta_2 \neq \delta_3$ when $\nu_1^X - \nu_2^X = 0.1, \alpha = 0.025, \nu_1^X - \nu_3^X = 0.3, \tau = 1, \sigma_X = \sigma_Y = 1, \epsilon = 0$ and $\rho = 0.5$ by $\delta_3$ and $n$ when different endpoints at interim and final analyses.

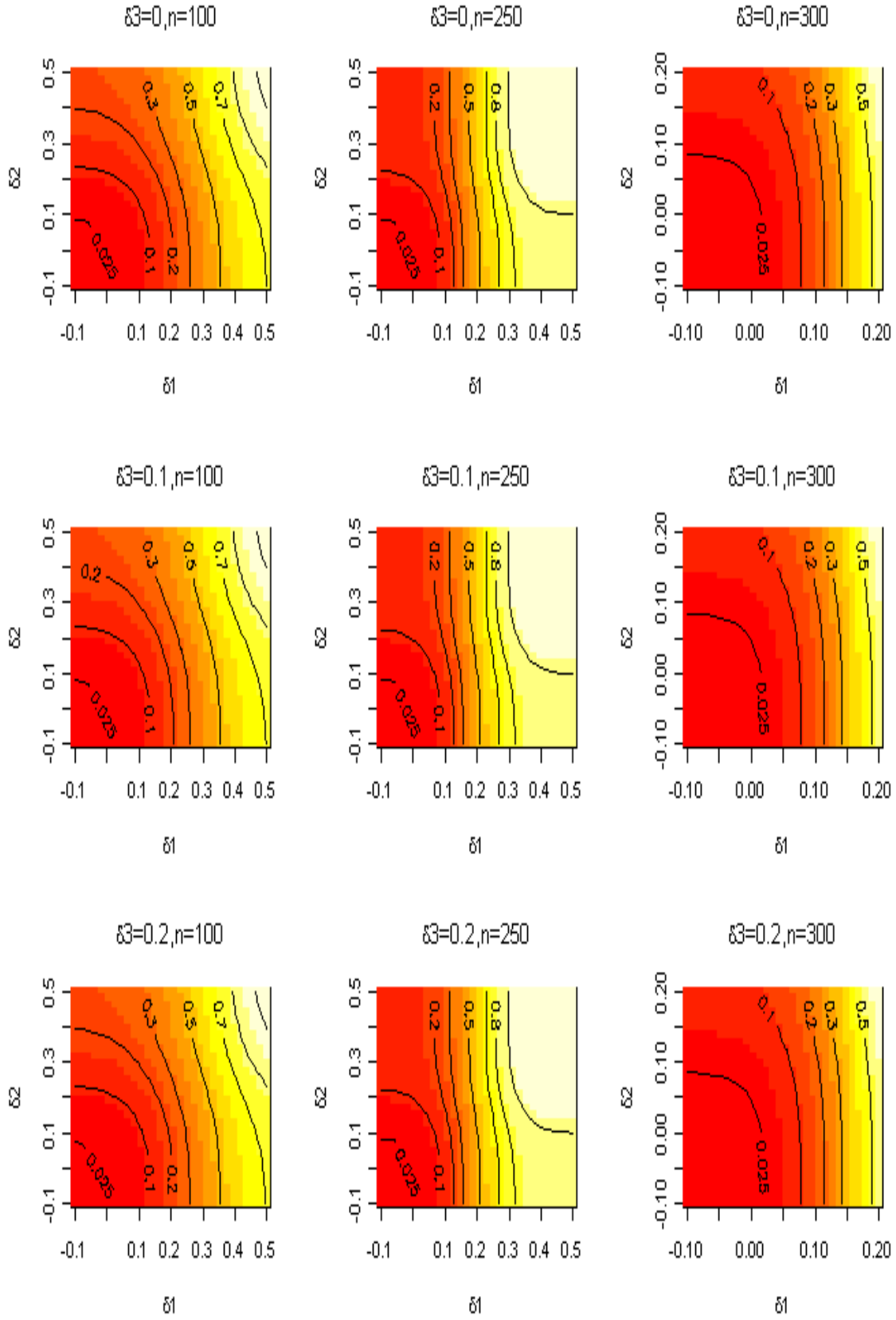


Figure 4.10: Contour of power surface for $\delta_1 \neq \delta_2 \neq \delta_3$ when $\nu_1^X - \nu_2^X = 0.1, \alpha = 0.025, \nu_1^X - \nu_3^X = 0.3, n = 250, \sigma_X = \sigma_Y = 1, \epsilon = 0$ and $\delta_3 = 0.1$ by $\tau$ and $\rho$ when different endpoints at interim and final analyses.

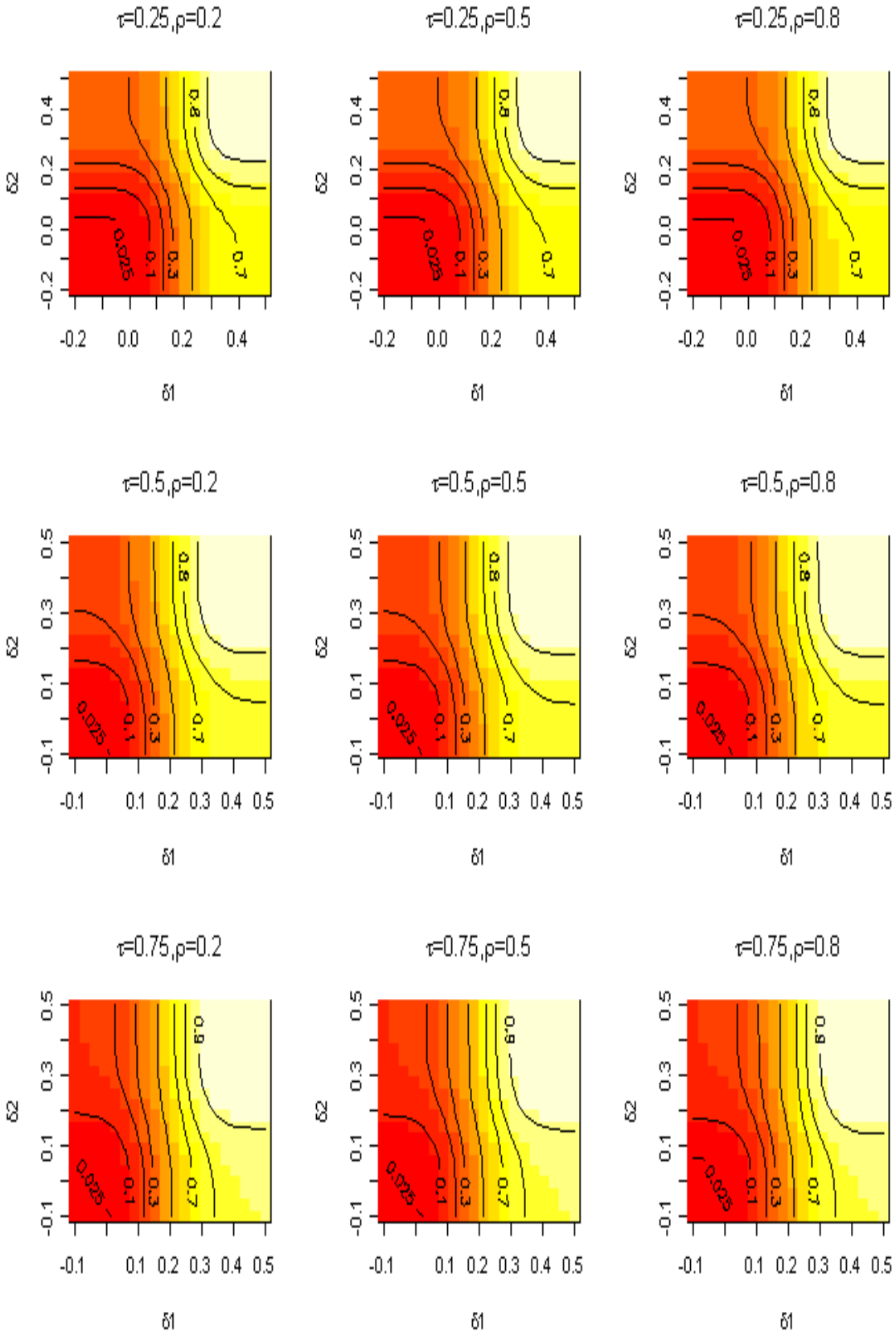


Figure 4.11: Contour of power surface for $\delta_1 \neq \delta_2 \neq \delta_3$ when $\nu_1^X - \nu_2^X = 0.1, \nu_1^X - \nu_3^X = 0.3, \alpha = 0.025, \tau = 1, \rho = 0.5 (i.e. \eta = 0.25), \sigma_X = \sigma_Y = 1$ and $\delta_3 = 0.1$ by $\rho$ and $\epsilon$ when different endpoints at interim and final analyses.

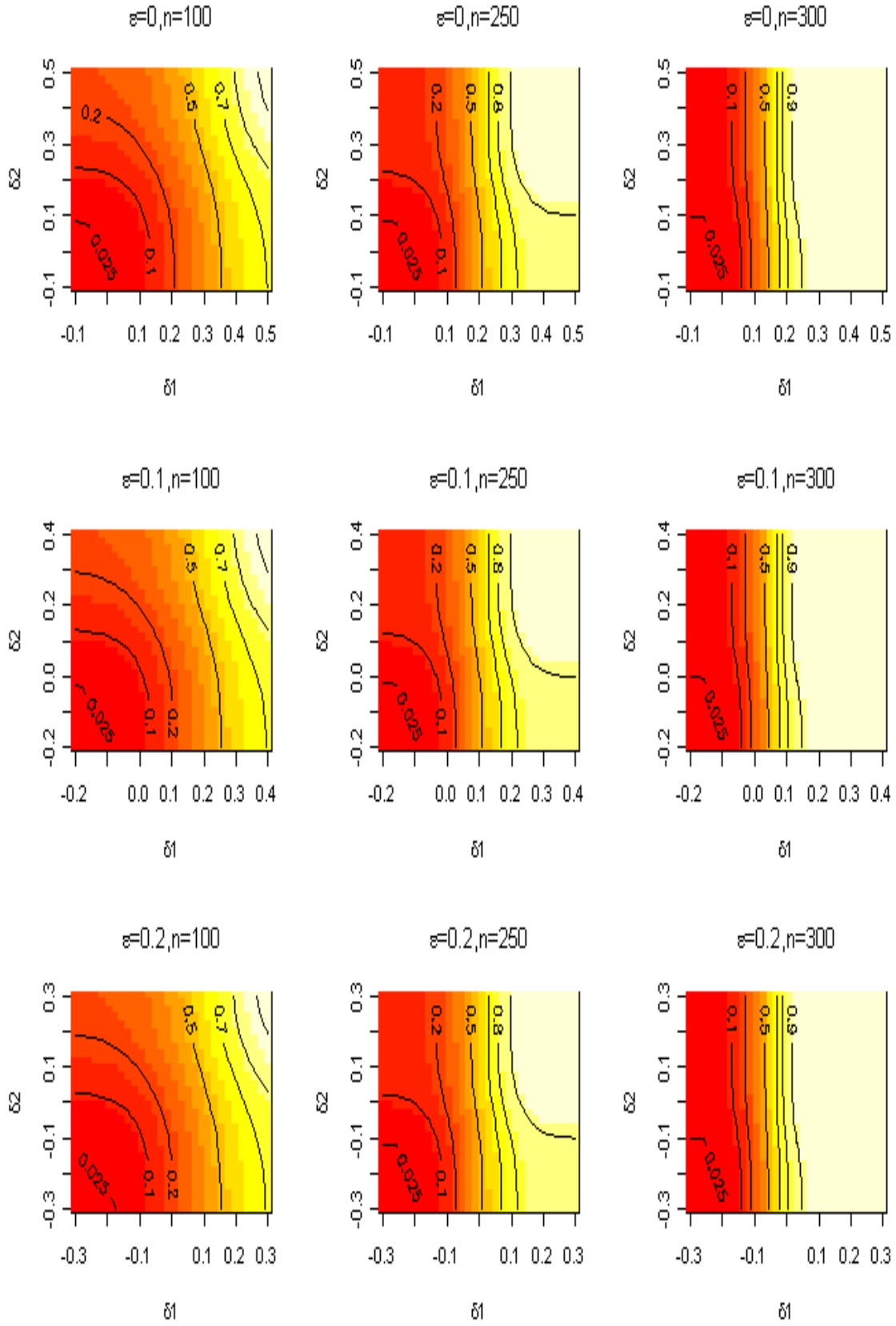
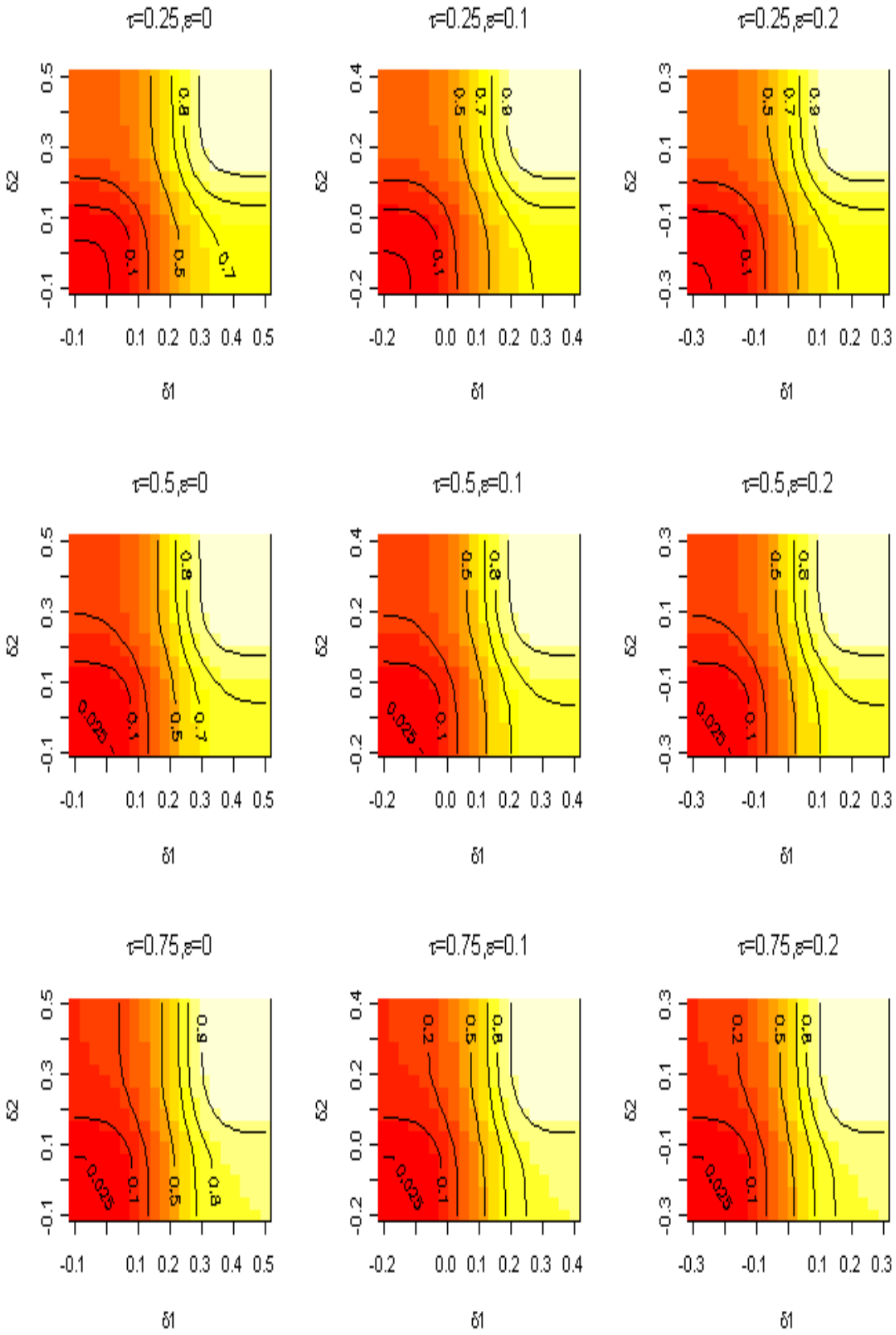


Figure 4.12: Contour of power surface for $\delta_1 \neq \delta_2 \neq \delta_3$ when $\nu_1^X - \nu_2^X = 0.1, \nu_1^X - \nu_3^X = 0.3, \alpha = 0.025, \delta_3 = 0.1, \rho = 0.8, \sigma_X = \sigma_Y = 1$ and $n = 250$ by $\tau$ and $\epsilon$ when different endpoints at interim and final analyses.



$$
\begin{aligned}
&-\sqrt{\frac{3}{4}-\eta^2}\int_{-\infty}^{c^*-w_2-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon}\phi(z)\int_{-\infty}^{\frac{\sqrt{\tau}(w_2-w_1)+\eta z}{\sqrt{1-\eta^2}}}\phi(s)\\
&\quad\times\Phi\left(\frac{(\sqrt{\tau}(w_2-w_3)+\eta z)\sqrt{1-\eta^2}-(\frac{1}{2}-\eta^2)s}{\sqrt{\frac{3}{4}-\eta^2}}\right)dsdz\\
&-\sqrt{\frac{3}{4}-\eta^2}\int_{-\infty}^{c^*-w_3-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon}\phi(z)\int_{-\infty}^{\frac{\sqrt{\tau}(w_3-w_1)+\eta z}{\sqrt{1-\eta^2}}}\phi(s)\\
&\quad\times\Phi\left(\frac{(\sqrt{\tau}(w_3-w_2)+\eta z)\sqrt{1-\eta^2}-(\frac{1}{2}-\eta^2)s}{\sqrt{\frac{3}{4}-\eta^2}}\right)dsdz. \qquad (4.16)
\end{aligned}
$$

When $\epsilon = 0$, Equation 4.16 becomes the power function for superiority hypothesis. Given type I error $\alpha$ and $\eta$, the critical value $c^*$ can be determined by Equation 4.10. In the following paragraphs, we discuss the power function under three scenarios.

- Case I: $\delta_1 = \delta_2 = \delta_3 = \delta$.

  When $\delta_1 = \delta_2 = \delta_3 = \delta$, the overall power $1-\beta$ can be calculated as:

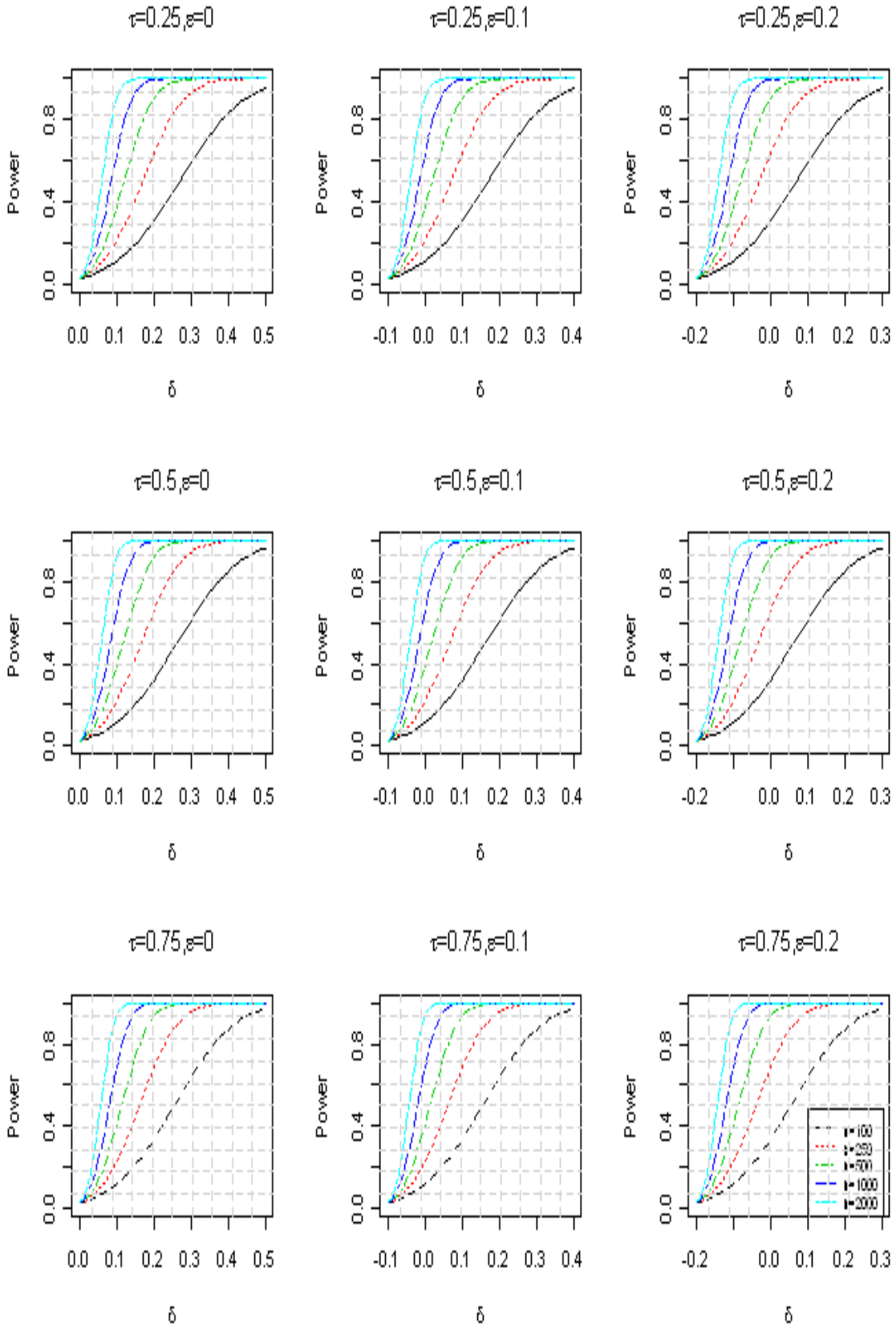
$$
\begin{aligned}
1-\beta &= Pr\left(W^* > c^*; w_1 = w_2 = w_3 = \sqrt{\frac{n}{2\sigma_Y^2}}\delta, \eta\right)\\
&=1-3\sqrt{\frac{3}{4}-\eta^2}\int_{-\infty}^{c^*-\sqrt{\frac{n}{2\sigma_Y^2}}(\delta+\epsilon)}\phi(z)\int_{-\infty}^{\frac{\eta z}{\sqrt{1-\eta^2}}}\phi(s)\Phi\left(\frac{\eta z\sqrt{1-\eta^2}-(\frac{1}{2}-\eta^2)s}{\sqrt{\frac{3}{4}-\eta^2}}\right)dsdz.
\end{aligned} \qquad (4.17)
$$

  From Equation 4.17, the power is a function of $c^*, \sigma_Y$, n, $\delta, \epsilon$ and $\eta$ (or $\tau$). Equation 4.17 shows that for a various combinations of $\delta$ and $\epsilon$, as long as the sum of these two parameters are the same, so is the power.

  Figure 4.13 shows the effects of $n, \tau$ and $\epsilon$ on the power curve for $\delta_1 = \delta_2 = \delta_3 = \delta$ when $\alpha = 0.025, \sigma_Y = 1$. As $\tau$ increases, there is no visual difference on the power, that is because now the interim and the final endpoints are the same, and under the case that $\delta_1 = \delta_2 = \delta_3$, there is no difference on the power regardless when or which treatment we select. On the other hand, power increases as $\epsilon$, $n$ or $\delta$ increases.

- Case II: $\delta_1 = \delta_2 = \delta \neq \delta_3$.

Figure 4.13: Power curve for $\delta_1 = \delta_2 = \delta_3 = \delta$ when $\alpha = 0.025, \sigma_Y = 1$ by information time $\tau$ and $\epsilon$ when same endpoints at interim and final analyses.



When $\delta_1 = \delta_2 = \delta \neq \delta_3$, the power $1-\beta$ can be calculated as:

$$\begin{aligned}
1-\beta =& 1 - \sqrt{\frac{3}{4}-\eta^2}\int_{-\infty}^{c^*-\sqrt{\frac{n}{2\sigma_Y^2}}(\delta+\epsilon)} \phi(z) \int_{-\infty}^{\frac{\eta z}{\sqrt{1-\eta^2}}} \phi(s) \\
&\times \Phi\Big(\frac{(\sqrt{\tau}(w_1-w_3)+\eta z)\sqrt{1-\eta^2} - (\frac{1}{2}-\eta^2)s}{\sqrt{\frac{3}{4}-\eta^2}}\Big) ds dz \\
&- \sqrt{\frac{3}{4}-\eta^2}\int_{-\infty}^{c^*-\sqrt{\frac{n}{2\sigma_Y^2}}(\delta+\epsilon)} \phi(z) \int_{-\infty}^{\frac{\eta z}{\sqrt{1-\eta^2}}} \phi(s) \\
&\times \Phi\Big(\frac{(\sqrt{\tau}(w_2-w_3)+\eta z)\sqrt{1-\eta^2} - (\frac{1}{2}-\eta^2)s}{\sqrt{\frac{3}{4}-\eta^2}}\Big) ds dz \\
&- \sqrt{\frac{3}{4}-\eta^2}\int_{-\infty}^{c^*-\sqrt{\frac{n}{2\sigma_Y^2}}(\delta_3+\epsilon)} \phi(z) \int_{-\infty}^{\frac{\sqrt{\tau}(w_3-w_1)+\eta z}{\sqrt{1-\eta^2}}} \phi(s) \\
&\times \Phi\Big(\frac{(\sqrt{\tau}(w_3-w_2)+\eta z)\sqrt{1-\eta^2} - (\frac{1}{2}-\eta^2)s}{\sqrt{\frac{3}{4}-\eta^2}}\Big) ds dz. \qquad (4.18)
\end{aligned}$$

From Equation 4.18, we know that the power is a function of $\tau, n, \sigma_Y, \delta_1 = \delta_2 = \delta, \delta_3$, and $\epsilon$.

Figure 4.14 shows the effects of $\tau$ and $\epsilon$ on the power when $\delta_1 = \delta_2 = \delta \neq \delta_3, \alpha = 0.025$ and $n = 250, \sigma_Y = 1$. Because the endpoints for the interim and the final analysis are the same, the later we perform the interim selection, the probability of selecting the right treatment in higher. This can been seen from the figures that the power surface shift to the left when the $\tau$ increases. The timing for the interim selection play an important role under this scenario.

Figure 4.15 shows the effects of $n$ and $\epsilon$ on the power for $\delta_1 = \delta_2 = \delta \neq \delta_3$ when $\alpha = 0.025, \sigma_Y = 1$ and $\tau = 0.5$. When sample size $n$ increases, the power increases. The power surface shift to the left when $\epsilon$ increase, this implies the power increase.

- Case III: $\delta_1 \neq \delta_2 \neq \delta_3$.

  When $\delta_1 \neq \delta_2 \neq \delta_3$, the power $1-\beta$ can be calculated as:

$$\begin{aligned}
1-\beta =& Pr\left(W^* > c^*; H_1\right) \\
=& \gamma(c^*; w_1, w_2, w_3, \eta)
\end{aligned}$$

Figure 4.14: Contour of power surface for $\delta_1 = \delta_2 = \delta \neq \delta_3$ when $n = 250, \sigma_Y = 1$ and $\alpha = 0.025$ by $\tau$ and $\epsilon$ when same endpoints at interim and final analyses.

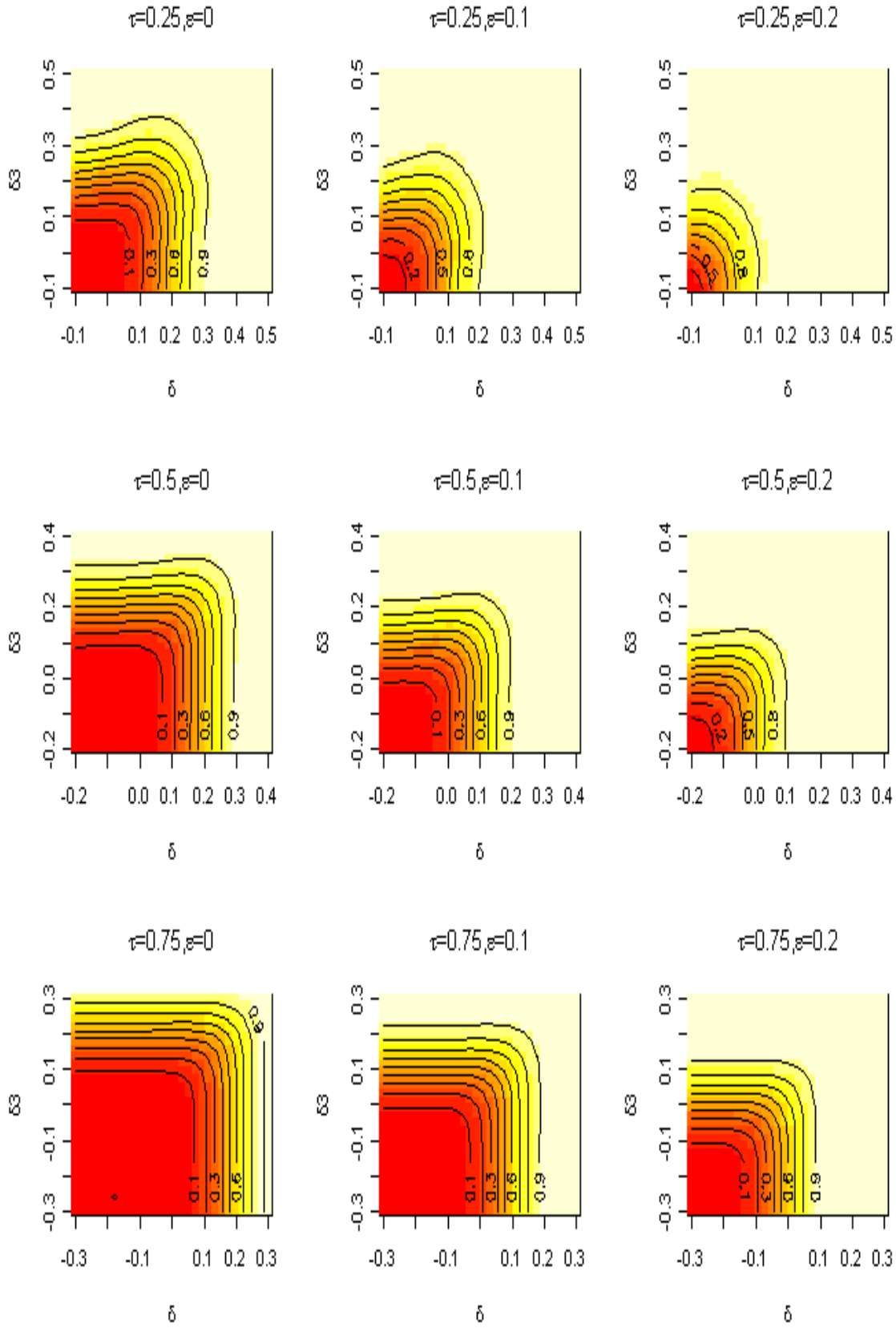
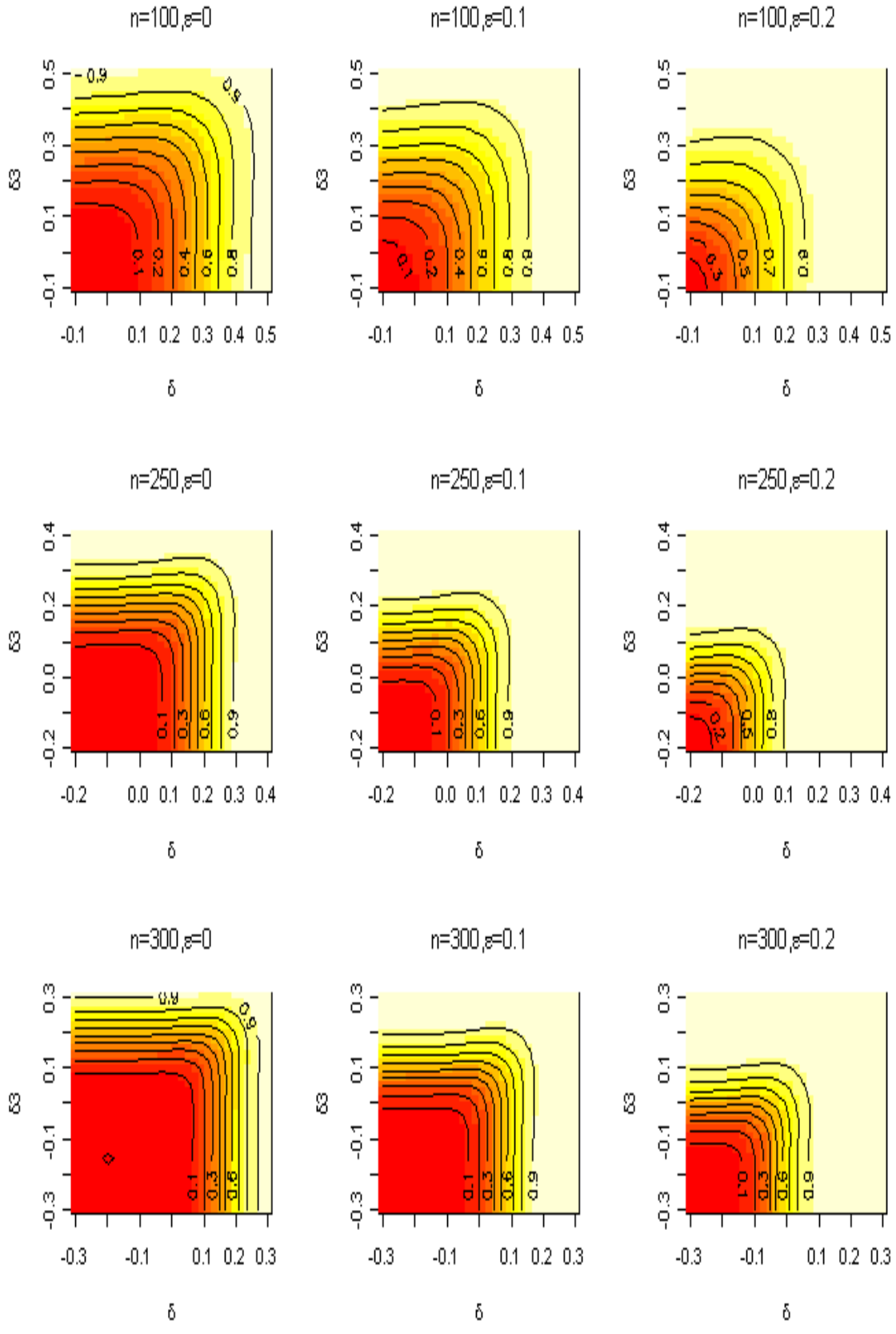


Figure 4.15: Contour of power surface for $\delta_1 = \delta_2 = \delta \neq \delta_3$ when $\tau = 0.5, \sigma_Y = 1$ and $\alpha = 0.025$ by $n$ and $\epsilon$ when same endpoints at interim and final analyses.



$$
\begin{aligned}
=&1-\sqrt{\frac{3}{4}-\eta^2}\int_{-\infty}^{c^*-\sqrt{\frac{n}{2\sigma_Y^2}}(\delta_1+\epsilon)}\phi(z)\int_{-\infty}^{\frac{\sqrt{\tau}(w_1-w_2)+\eta z}{\sqrt{1-\eta^2}}}\phi(s)\\
&\times\Phi(\frac{(\sqrt{\tau}(w_1-w_3)+\eta z)\sqrt{1-\eta^2}-(\frac{1}{2}-\eta^2)s}{\sqrt{\frac{3}{4}-\eta^2}})dsdz\\
&-\sqrt{\frac{3}{4}-\eta^2}\int_{-\infty}^{c^*-\sqrt{\frac{n}{2\sigma_Y^2}}(\delta_2+\epsilon)}\phi(z)\int_{-\infty}^{\frac{\sqrt{\tau}(w_2-w_1)+\eta z}{\sqrt{1-\eta^2}}}\phi(s)\\
&\times\Phi(\frac{(\sqrt{\tau}(w_2-w_3)+\eta z)\sqrt{1-\eta^2}-(\frac{1}{2}-\eta^2)s}{\sqrt{\frac{3}{4}-\eta^2}})dsdz\\
&-\sqrt{\frac{3}{4}-\eta^2}\int_{-\infty}^{c^*-\sqrt{\frac{n}{2\sigma_Y^2}}(\delta_2+\epsilon)}\phi(z)\int_{-\infty}^{\frac{\sqrt{\tau}(w_3-w_1)+\eta z}{\sqrt{1-\eta^2}}}\phi(s)\\
&\times\Phi(\frac{(\sqrt{\tau}(w_3-w_2)+\eta z)\sqrt{1-\eta^2}-(\frac{1}{2}-\eta^2)s}{\sqrt{\frac{3}{4}-\eta^2}})dsdz. \qquad (4.19)
\end{aligned}
$$

From Equation 4.19, the power is a function of $\tau, n, \delta_1, \delta_2, \delta_3, \epsilon$.

Figure 4.16 shows the effects of $\delta_3$ and $\epsilon$ on the power for $\delta_1 \neq \delta_2 \neq \delta_3$ when $\tau = 0.5, n = 250, \sigma_Y = 1$ and $\alpha = 0.025$. When $\delta_3$ increases, the power increases. As seen in the figures, the power surface is not monotonic with respective to $\delta_1$ or $\delta_2$ or $\delta_3$.

Figure 4.17 shows effects of $\tau$ and $\epsilon$ on power for $\delta_1 \neq \delta_2 \neq \delta_3$ when $\delta_3 = 0.1, n = 250, \sigma_Y = 1$ and $\alpha = 0.025$. When $\tau$ increases, the power increases. Notice that the power surface is not monotonic with respective to $\delta_1$ or $\delta_2$ or $\delta_3$.

Figure 4.18 shows the effects of $n$ and $\epsilon$ on power for $\delta_1 \neq \delta_2 \neq \delta_3$ when $\delta_3 = 0.1, \tau = 0.5, \sigma_Y = 1$ and $\alpha = 0.025$. When $n$ increases, the power increases. When $\epsilon$ increase to 0.1 or 0.2, or when sample size increase to 300, the power surface is not monotonic with respective to $\delta_1$ or $\delta_2$ or $\delta_3$.

## 4.5 Sample Size Estimation

### 4.5.1 Different Endpoints at Interim and Final Analyses

- Case I: $\delta_1 = \delta_2 = \delta_3 = \delta$.

  When $\delta_1 = \delta_2 = \delta_3 = \delta$, as shown in Equation 4.13, sample size can be determined

Figure 4.16: Contour of power surface for $\delta_1 \neq \delta_2 \neq \delta_3$ when $\tau = 0.5, n = 250, \sigma_Y = 1$ and $\alpha = 0.025$ by $\delta_3$ and $\epsilon$ when same endpoints at interim and final analyses.

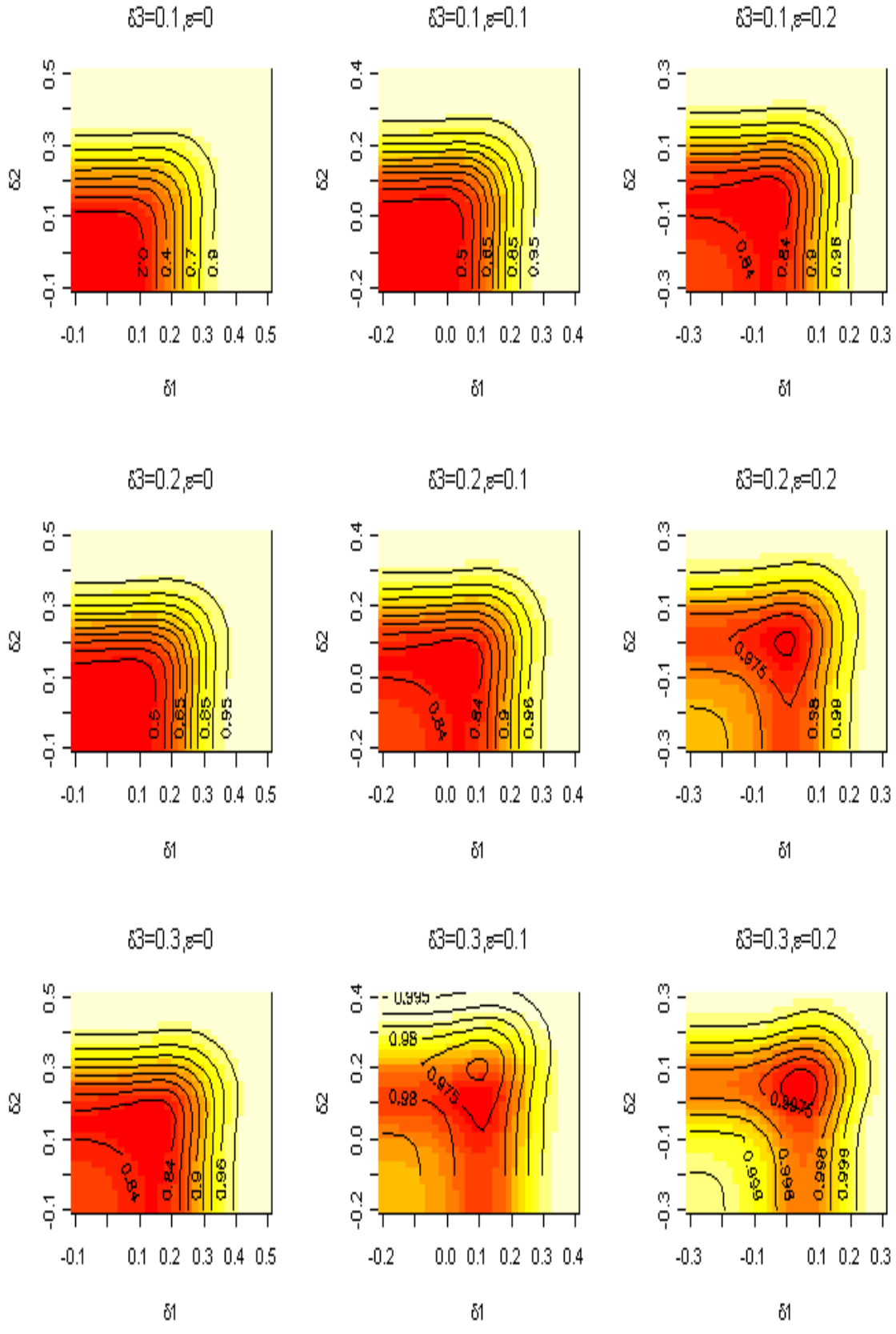


Figure 4.17: Contour of power surface when $\delta_1 \neq \delta_2 \neq \delta_3$ with vary information time $\tau$ and $\epsilon$ when $\delta_3 = 0.1, n = 250, \sigma_Y = 1$ and $\alpha = 0.025$ when same endpoints at interim and final analyses.

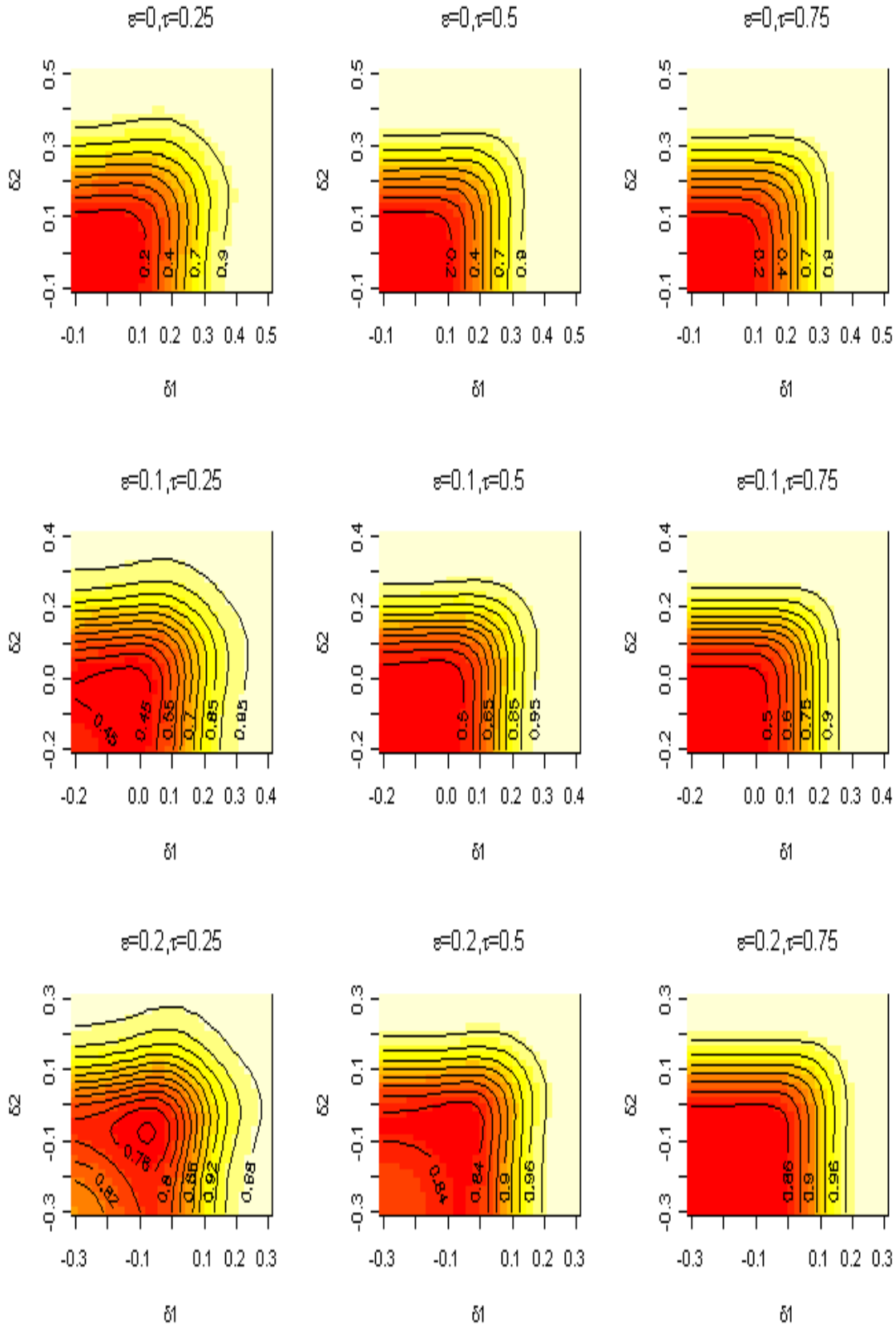


Figure 4.18: Contour of power surface when $\delta_1 \neq \delta_2 \neq \delta_3$ by n and $\epsilon$ when $\tau = 0.5, \delta_3 = 0.1, \sigma_Y = 1$ and $\alpha = 0.025$ when same endpoints at interim and final analyses.

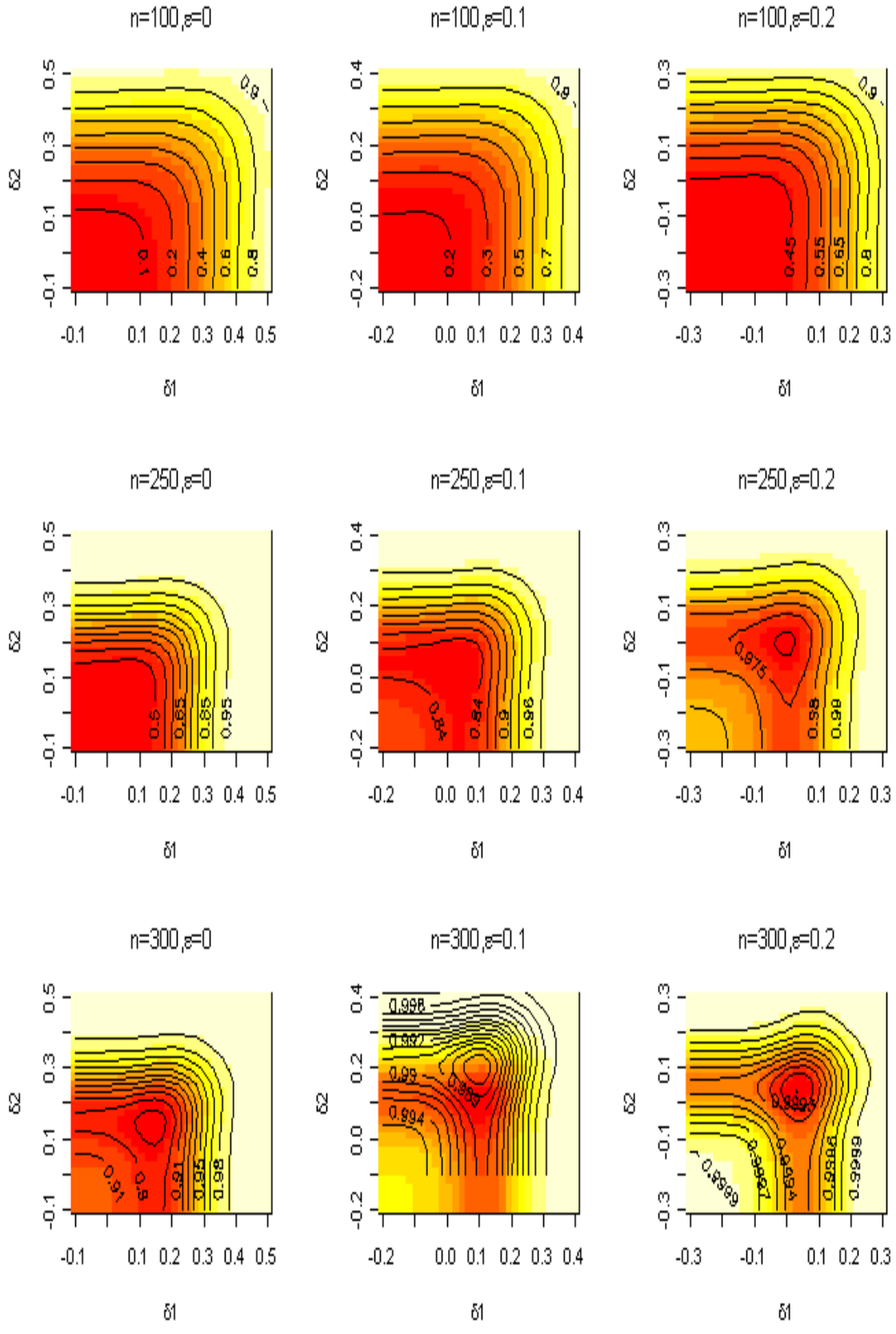


Table 4.4: Estimated sample size for $\delta_1 = \delta_2 = \delta_3 = \delta$, when $\nu_1^X - \nu_2^X = 0.1$, $\nu_1^X - \nu_3^X = 0.3, \sigma_Y = \sigma_X = 1, \alpha = 0.025$ and $\epsilon = 0$ with targeted power of 80% or 90% by $\tau, \rho$ and $\delta$ when different endpoints used at the interim and the final analysis.

| | $\tau = 0.25$ | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\rho = 0.2$ | | | | | $\rho = 0.5$ | | | | | $\rho = 0.8$ | | | | |
| $\delta$ | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 |
| $1-\beta = 0.8$ | 1619 | 400 | 177 | 99 | 64 | 1686 | 408 | 178 | 100 | 63 | 1744 | 413 | 179 | 99 | 63 |
| $1-\beta = 0.9$ | 2164 | 535 | 237 | 133 | 85 | 2249 | 546 | 239 | 133 | 85 | 2324 | 553 | 240 | 133 | 84 |
| | $\tau = 0.5$ | | | | | | | | | | | | | | |
| | $\rho = 0.2$ | | | | | $\rho = 0.5$ | | | | | $\rho = 0.8$ | | | | |
| $\delta$ | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 |
| $1-\beta = 0.8$ | 1653 | 406 | 179 | 100 | 64 | 1766 | 421 | 183 | 101 | 64 | 1862 | 431 | 183 | 100 | 63 |
| $1-\beta = 0.9$ | 2203 | 543 | 239 | 134 | 86 | 2338 | 563 | 244 | 135 | 86 | 2453 | 577 | 246 | 135 | 85 |
| | $\tau = 0.75$ | | | | | | | | | | | | | | |
| | $\rho = 0.2$ | | | | | $\rho = 0.5$ | | | | | $\rho = 0.8$ | | | | |
| $\delta$ | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 |
| $1-\beta = 0.8$ | 1678 | 411 | 181 | 101 | 64 | 1819 | 433 | 186 | 103 | 65 | 1933 | 448 | 188 | 102 | 63 |
| $1-\beta = 0.9$ | 2229 | 550 | 241 | 135 | 86 | 2393 | 579 | 249 | 137 | 87 | 2525 | 599 | 252 | 137 | 85 |

given the type I error $\alpha$, power $1-\beta$, $\eta$, $c^*, \delta, \lambda^{(12)}, \lambda^{(13)}, \sigma_Y$ and $\epsilon$.

Table 4.4 provides an example of the estimated sample size for $\delta_1 = \delta_2 = \delta_3 = \delta$, when $\nu_1^X - \nu_2^X = 0.1$, $\nu_1^X - \nu_3^X = 0.3, \sigma_Y = \sigma_X = 1, \alpha = 0.025$ and $\epsilon = 0$ with targeted power of 80% or 90% by $\tau, \rho$ and $\delta$. At a given $\tau$, $\delta$ and power, as $\rho$ increases, the sample size increases. Given power, when $\delta$ increases, the required sample size decreases and when $\tau$ increases, sample size increases.

Table 4.5 provides an example of the estimated sample size for $\delta_1 = \delta_2 = \delta_3 = \delta$, when $\nu_1^X - \nu_2^X = 0.1$, $\nu_1^X - \nu_3^X = 0.3, \sigma_Y = \sigma_X = 1, \tau = 0.5$ and $\alpha = 0.025$ with targeted power of 80% or 90% by $\rho, \epsilon$ and $\delta$. From Equation 4.13, we know that as long as the sum of the $\delta$ and the $\epsilon$ are the same, sample size are the same when the other parameters are fixed. For example, in table 4.4, when $\rho$ is fixed, the estimated sample size for $\delta = 0.1, \epsilon = 0.2$ is the same as for $\delta = 0.2, \epsilon = 0.1$. When $\delta$ increases from 0.1 to 0.2, the required sample size drop dramatically. When $\epsilon = 0$, the sample size will matches the sample size estimated in table 4.4 with the same parameters.

- Case II: $\delta_1 = \delta_2 = \delta \neq \delta_3$.

  When $\delta_1 = \delta_2 = \delta \neq \delta_3$, as shown in Equation 4.14, sample size can be determined given the type I error $\alpha$, power $1-\beta$, $\eta$, $c^*, \delta, \delta_3, \lambda^{(12)}, \lambda^{(13)}, \sigma_Y$ and $\epsilon$.

Table 4.5: Estimated sample size for $\delta_1 = \delta_2 = \delta_3 = \delta$, when $\nu_1^X - \nu_2^X = 0.1$, $\nu_1^X - \nu_3^X = 0.3, \sigma_Y = \sigma_X = 1, \tau = 0.5$ and $\alpha = 0.025$ with targeted power of 80% or 90% by $\rho, \epsilon$ and $\delta$ when different endpoints used at the interim and the final analysis.

| | $\epsilon = 0$ | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\rho = 0.2$ | | | | | $\rho = 0.5$ | | | | | $\rho = 0.8$ | | | | |
| $\delta$ | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 |
| $1-\beta = 0.8$ | 1653 | 406 | 179 | 100 | 64 | 1766 | 421 | 183 | 101 | 64 | 1862 | 431 | 183 | 100 | 63 |
| $1-\beta = 0.9$ | 2203 | 543 | 239 | 134 | 86 | 2338 | 563 | 244 | 135 | 86 | 2453 | 577 | 246 | 135 | 85 |
| | $\epsilon = 0.1$ | | | | | | | | | | | | | | |
| | $\rho = 0.2$ | | | | | $\rho = 0.5$ | | | | | $\rho = 0.8$ | | | | |
| $\delta$ | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 |
| $1-\beta = 0.8$ | 406 | 179 | 100 | 64 | 44 | 421 | 183 | 101 | 64 | 44 | 431 | 183 | 100 | 63 | 43 |
| $1-\beta = 0.9$ | 543 | 239 | 134 | 86 | 59 | 563 | 244 | 135 | 86 | 59 | 577 | 246 | 135 | 85 | 58 |
| | $\epsilon = 0.2$ | | | | | | | | | | | | | | |
| | $\rho = 0.2$ | | | | | $\rho = 0.5$ | | | | | $\rho = 0.8$ | | | | |
| $\delta$ | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 |
| $1-\beta = 0.8$ | 179 | 100 | 64 | 44 | 33 | 183 | 101 | 64 | 44 | 32 | 183 | 100 | 63 | 43 | 32 |
| $1-\beta = 0.9$ | 239 | 134 | 86 | 59 | 44 | 244 | 135 | 86 | 59 | 43 | 246 | 135 | 85 | 58 | 42 |

Table 4.6 provides an example of the estimated sample size for $\delta_1 = \delta_2 = \delta \neq \delta_3$ when $\nu_1^X - \nu_2^X = 0.1$, $\nu_1^X - \nu_3^X = 0.3, \sigma_Y = \sigma_X = 1, \rho = 0.8$ and $\alpha = 0.025$ with targeted power of 80% or 90% by $\tau, \epsilon, \delta$ and $\delta_3$. When the two treatment are the same and close to the third treatment (for example $\delta_1 = \delta_2 = 0.5 > \delta_3 = 0.4$), we will need a smaller sample size to achieve the targeted power compare to when two treatment are the same and far apart from the third treatment arm (for example $\delta_1 = \delta_2 = 0.5 > \delta_3 = 0.1$). When the non-inferiority margin increase, a smaller sample will be needed to achieve the targeted power.

- Case III: $\delta_1 \neq \delta_2 \neq \delta_3$.

  When $\delta_1 \neq \delta_2 \neq \delta_3$, as shown in Equation 4.15, sample size can be determined given the type I error $\alpha$, power $1-\beta$, $\eta$, $c^*, \delta_1, \delta_2, \delta_3, \lambda^{(12)}, \lambda^{(13)}, \sigma_Y$ and $\epsilon$.

  Table 4.7 provides an example of the estimated sample size for $\delta_1 \neq \delta_2 \neq \delta_3$ when $\nu_1^X - \nu_2^X = 0.1$, $\nu_1^X - \nu_3^X = 0.3, \sigma_Y = \sigma_X = 1, \rho = 0.8$ and $\alpha = 0.025$ with targeted power of 80% or 90% by $\tau, \epsilon, \delta_1, \delta_2$ and $\delta_3$. When the treatment effects for the surrogate endpoint are in the same direction as the final endpoints, (i.e. $\nu_{12} > \nu_{13}$ is in the same direction as $\delta_1 < \delta_2 < \delta_3$), performing the interim analysis at a later time, it will save the sample size to maintain the same power. For example, in table 4.7, we choose

Table 4.6: Estimated sample size for $\delta_1 = \delta_2 = \delta \neq \delta_3$ when $\nu_1^X - \nu_2^X = 0.1$, $\nu_1^X - \nu_3^X = 0.3, \sigma_Y = \sigma_X = 1, \rho = 0.8$ and $\alpha = 0.025$ with targeted power of 80% or 90% by $\tau, \epsilon, \delta$ and $\delta_3$ when different endpoints used at the interim and the final analysis.

| | $\tau = 0.25, \rho = 0.8, (i.e. \eta = 0.2)$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | $\epsilon = 0$ | | | | $\epsilon = 0.1$ | | | |
| $(\delta, \delta_3)$ | (0.1,0.3) | (0.3,0.4) | (0.5,0.4) | (0.5,0.1) | (0.1,0.3) | (0.3,0.4) | (0.5,0.4) | (0.5,0.1) |
| $1 - \beta = 0.8$ | 1744 | 177 | 66 | 78 | 412 | 97 | 46 | 55 |
| $1 - \beta = 0.9$ | 2324 | 239 | 88 | 109 | 553 | 131 | 61 | 79 |
| | $\tau = 0.5, \rho = 0.8, (i.e. \eta = 0.28)$ | | | | | | | |
| | $\epsilon = 0$ | | | | $\epsilon = 0.1$ | | | |
| $(\delta, \delta_3)$ | (0.1,0.3) | (0.3,0.4) | (0.5,0.4) | (0.5,0.1) | (0.1,0.3) | (0.3,0.4) | (0.5,0.4) | (0.5,0.1) |
| $1 - \beta = 0.8$ | 1862 | 183 | 64 | 71 | 431 | 100 | 44 | 50 |
| $1 - \beta = 0.9$ | 2453 | 246 | 86 | 95 | 577 | 135 | 59 | 67 |
| | $\tau = 0.75, \rho = 0.8, (i.e. \eta = 0.35)$ | | | | | | | |
| | $\epsilon = 0$ | | | | $\epsilon = 0.1$ | | | |
| $(\delta, \delta_3)$ | (0.1,0.3) | (0.3,0.4) | (0.5,0.4) | (0.5,0.1) | (0.1,0.3) | (0.3,0.4) | (0.5,0.4) | (0.5,0.1) |
| $1 - \beta = 0.8$ | 1933 | 188 | 64 | 68 | 448 | 102 | 44 | 47 |
| $1 - \beta = 0.9$ | 2525 | 252 | 86 | 90 | 599 | 137 | 59 | 63 |

$\nu_{12} = 0.1 < \nu_{13} = 0.3$, which is in the same direction as final endpoint $\delta_1 > \delta_2 > \delta_3$. When $\epsilon = 0$ and $\delta_1 = 0.3 > \delta_2 = 0.2 > \delta_3 = 0.1$, performing the interim analysis at $\tau = 0.25$, we need 242 subjects to achieve 80% power, whereas when we perform interim selection at $\tau = 0.5$, we only need 224 subjects to achieve 80% power. If the treatment effect of the final endpoint are in the opposite direction as the surrogate endpoints, we need a larger sample size to achieve the targeted power. For example, in table 4.7, when $\delta_1 = 0.1 < \delta_2 = 0.2 < \delta_3 = 0.3$, the sample size required to achieve 80% power is 1705, on the other hand, when $\delta_1 = 0.3 > \delta_2 = 0.2 > \delta_3 = 0.2$, we only need 242 subjects to achieve the same power. When the non-inferiority margin increase, smaller sample is needed to achieve the targeted power.

### 4.5.2 Same Endpoints at Interim and Final Analyses

- Case I: $\delta_1 = \delta_2 = \delta_3 = \delta$.

  When $\delta_1 = \delta_2 = \delta_3 = \delta$, as shown in Equation 4.17, sample size can be determined given the type I error $\alpha$, power $1 - \beta$, $\tau$ (or $\eta$), $c^*, \delta, \sigma_Y$ and $\epsilon$.

  Table 4.8 provides an example of the estimated sample size for $\delta_1 = \delta_2 = \delta_3 = \delta$ when $\sigma_Y = 1$ and $\alpha = 0.025$ with targeted power of 80% or 90% by $\tau, \epsilon$ and $\delta$. From

Table 4.7: Estimated sample size for $\delta_1 \neq \delta_2 \neq \delta_3$ when $\nu_1^X - \nu_2^X = 0.1$, $\nu_1^X - \nu_3^X = 0.3, \sigma_Y = \sigma_X = 1, \rho = 0.8$ and $\alpha = 0.025$ with targeted power of 80% or 90% by $\tau, \epsilon, \delta_1, \delta_2$ and $\delta_3$ when different endpoints used at the interim and the final analysis.

| | $\tau = 0.25, \rho = 0.8 (i.e. \eta = 0.2)$ | | | | | |
|---|---|---|---|---|---|---|
| | $\epsilon = 0$ | | | $\epsilon = 0.1$ | | |
| $(\delta_1, \delta_2, \delta_3)$ | (0.1,0.2,0.3) | (0.2,0.3,0.5) | (0.3,0.2,0.1) | (0.1,0.2,0.3) | (0.2,0.3,0.5) | (0.3,0.2,0.1) |
| $1 - \beta = 0.8$ | 1705 | 366 | 242 | 366 | 151 | 131 |
| $1 - \beta = 0.9$ | 2308 | 520 | 328 | 520 | 216 | 177 |
| | $\tau = 0.5, \rho = 0.8 (i.e. \eta = 0.28)$ | | | | | |
| | $\epsilon = 0$ | | | $\epsilon = 0.1$ | | |
| $(\delta_1, \delta_2, \delta_3)$ | (0.1,0.2,0.3) | (0.2,0.3,0.5) | (0.3,0.2,0.1) | (0.1,0.2,0.3) | (0.2,0.3,0.5) | (0.3,0.2,0.1) |
| $1 - \beta = 0.8$ | 1858 | 409 | 224 | 409 | 166 | 123 |
| $1 - \beta = 0.9$ | 2452 | 566 | 299 | 566 | 233 | 163 |
| | $\tau = 0.75, \rho = 0.8 (i.e. \eta = 0.35)$ | | | | | |
| | $\epsilon = 0$ | | | $\epsilon = 0.1$ | | |
| $(\delta_1, \delta_2, \delta_3)$ | (0.1,0.2,0.3) | (0.2,0.3,0.5) | (0.3,0.2,0.1) | (0.1,0.2,0.3) | (0.2,0.3,0.5) | (0.3,0.2,0.1) |
| $1 - \beta = 0.8$ | 1932 | 438 | 218 | 438 | 176 | 119 |
| $1 - \beta = 0.9$ | 2525 | 595 | 287 | 595 | 245 | 157 |

Equation 4.17, we know that as long as the sum of the $\delta$ and the $\epsilon$ are the same, sample size are the same when the other parameters are fixed. For example, in table 4.8, when $\tau = 0.25, \epsilon = 0$ and $\delta = 0.3$, the estimated sample size needed to achieve 80% power is 166, which is the same then when $\tau = 0.25, \epsilon = 0.1$ and $\delta = 0.2$. When $\delta$ increases from 0.1 to 0.2, the required sample size drop dramatically. When same endpoints are used at the interim and the final analysis, when we perform the interim selection at a later time, we need a smaller sample size to achieve the targeted power. For example, in table 4.8, when $\epsilon = 0$, and $\delta = 0.1$, as $\tau$ increases from 0.25 to 0.5, the sample size required to achieve 80% power drop from 1486 to 1404.

- Case II: $\delta_1 = \delta_2 = \delta \neq \delta_3$.

  When $\delta_1 = \delta_2 = \delta \neq \delta_3$, as shown in Equation 4.18, sample size can be determined given the type I error $\alpha$, power $1 - \beta$, $\tau$ (or $\eta$), $c^*, \delta, \delta_3, \sigma_Y$ and $\epsilon$.

  Table 4.9 provides an example of the estimated sample size for $\delta_1 = \delta_2 = \delta \neq \delta_3$ when $\sigma_Y = 1$ and $\alpha = 0.025$ with targeted power of 80% or 90% by $\tau, \epsilon, \delta$ and $\delta_3$. If the two experimental treatments that have the same effect is selected at interim analysis and is inferior to the third arm (i.e. $\delta_1 = \delta_2 < \delta_3$), we need to have a larger sample size to reach the targeted power compare to when the two is superior than

Table 4.8: Estimated sample size for $\delta_1 = \delta_2 = \delta_3 = \delta$ when $\sigma_Y = 1$ and $\alpha = 0.025$ with targeted power of 80% or 90% by $\tau, \epsilon$ and $\delta$ when same endpoints used at the interim and the final analysis.

| | $\tau = 0.25$ | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\epsilon = 0$ | | | | | $\epsilon = 0.1$ | | | | | $\epsilon = 0.2$ | | | | |
| $\delta$ | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 |
| $1-\beta = 0.8$ | 1486 | 372 | 166 | 93 | 60 | 372 | 166 | 93 | 60 | 42 | 166 | 93 | 60 | 42 | 31 |
| $1-\beta = 0.9$ | 1988 | 497 | 221 | 125 | 80 | 497 | 211 | 125 | 80 | 56 | 221 | 125 | 80 | 56 | 40 |
| | $\tau = 0.5$ | | | | | | | | | | | | | | |
| | $\epsilon = 0$ | | | | | $\epsilon = 0.1$ | | | | | $\epsilon = 0.2$ | | | | |
| $\delta$ | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 |
| $1-\beta = 0.8$ | 1404 | 351 | 156 | 88 | 57 | 351 | 156 | 88 | 57 | 39 | 156 | 88 | 57 | 39 | 29 |
| $1-\beta = 0.9$ | 1876 | 469 | 209 | 118 | 76 | 469 | 209 | 118 | 76 | 53 | 209 | 118 | 76 | 53 | 39 |
| | $\tau = 0.75$ | | | | | | | | | | | | | | |
| | $\epsilon = 0$ | | | | | $\epsilon = 0.1$ | | | | | $\epsilon = 0.2$ | | | | |
| $\delta$ | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 |
| $1-\beta = 0.8$ | 1324 | 331 | 148 | 83 | 53 | 331 | 148 | 83 | 53 | 37 | 148 | 83 | 53 | 37 | 28 |
| $1-\beta = 0.9$ | 1765 | 442 | 197 | 111 | 71 | 442 | 197 | 111 | 71 | 50 | 197 | 111 | 71 | 50 | 37 |

the third arm (i.e. $\delta_1 = \delta_2 > \delta_3$). For example, in table 4.9, at $\tau = 0.25, \epsilon = 0$, when $\delta_1 = \delta_2 = 0.1 < \delta_3 = 0.2$, we need 623 subjects to achieve 80% power. Whereas, when $\delta_1 = \delta_2 = 0.2 > \delta_3 = 0.1$, we only need 442 subjects to achieve 80% power. The sample size needed is smaller as the treatment are far apart than the treatment are close. For example, in table 4.9, at $\tau = 0.25, \epsilon = 0$, when $\delta_1 = \delta_2 = 0.1 < \delta_3 = 0.2$, we need at least 623 samples to achieve 80% power. On the other hand, when $\delta_1 = \delta_2 = 0.1 < \delta_3 = 0.5$, we only need 96 subjects to achieve 80% power. As $\tau$ or $\epsilon$ increases, the required sample size to reach the targeted power decreases.

- Case III: $\delta_1 \neq \delta_2 \neq \delta_3$.

  When $\delta_1 \neq \delta_2 \neq \delta_3$, as shown in Equation 4.19, sample size can be determined given the type I error $\alpha$, power $1-\beta$, $\tau$ (or $\eta$), $c^*, \delta_1, \delta_2, \delta_3, \sigma_Y$ and $\epsilon$. From Equation 4.19, we can see that as long as the value of the $\delta$'s are the same, regardless which one is superior or inferior (i.e. $\delta_1 > \delta_2 > \delta_3$ or $\delta_1 < \delta_2 < \delta_3$), the sample size needed will be the same by fixing other parameters. For example, in table 4.10, at $\tau = 0.25, \epsilon = 0$, when $\delta_1 = 0.1 < \delta_2 = 0.2 < \delta_3 = 0.3$, we need 259 subjects to achieve 80 power, which is the same when $\delta_1 = 0.3 > \delta_2 = 0.2 > \delta_3 = 0.1$ (results are not shown in the table).

  Table 4.10 provides an example of the estimated sample size for $\delta_1 \neq \delta_2 \neq \delta_3$

Table 4.9: Estimated sample size for $\delta_1 = \delta_2 = \delta \neq \delta_3$ when $\sigma_Y = 1$ and $\alpha = 0.025$ with targeted power of 80% or 90% by $\tau, \epsilon, \delta$ and $\delta_3$ when same endpoints used at the interim and the final analysis.

| | $\tau = 0.25$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | $\epsilon = 0$ | | | | $\epsilon = 0.1$ | | | |
| $(\delta, \delta_3)$ | (0.1,0.2) | (0.1,0.5) | (0.2,0.1) | (0.5,0.1) | (0.1,0.2) | (0.1,0.5) | (0.2,0.1) | (0.5,0.1) |
| $1-\beta = 0.8$ | 623 | 96 | 442 | 69 | 247 | 70 | 193 | 49 |
| $1-\beta = 0.9$ | 877 | 132 | 598 | 91 | 338 | 99 | 261 | 65 |
| | $\tau = 0.5$ | | | | | | | |
| | $\epsilon = 0$ | | | | $\epsilon = 0.1$ | | | |
| $(\delta, \delta_3)$ | (0.1,0.2) | (0.1,0.5) | (0.2,0.1) | (0.5,0.1) | (0.1,0.2) | (0.1,0.5) | (0.2,0.1) | (0.5,0.1) |
| $1-\beta = 0.8$ | 532 | 83 | 405 | 64 | 222 | 59 | 179 | 45 |
| $1-\beta = 0.9$ | 709 | 108 | 537 | 85 | 298 | 78 | 238 | 59 |
| | $\tau = 0.75$ | | | | | | | |
| | $\epsilon = 0$ | | | | $\epsilon = 0.1$ | | | |
| $(\delta, \delta_3)$ | (0.1,0.2) | (0.1,0.5) | (0.2,0.1) | (0.5,0.1) | (0.1,0.2) | (0.1,0.5) | (0.2,0.1) | (0.5,0.1) |
| $1-\beta = 0.8$ | 499 | 81 | 384 | 62 | 208 | 57 | 168 | 43 |
| $1-\beta = 0.9$ | 655 | 105 | 507 | 81 | 275 | 74 | 223 | 57 |

when $\sigma_Y = 1$ and $\alpha = 0.025$ with targeted power of 80% or 90% by $\tau, \epsilon, \delta_1, \delta_2$ and $\delta_3$. As $\tau$ increases, by fixing the other parameters, the required sample size to achieve the targeted power decreases. For example, in table 4.10, at $\epsilon = 0$, $(\delta_1 = 0.1, \delta_2 = 0.3, \delta_3 = 0.5)$, we need 95 subjects to achieve 80% power when we perform interim selection at $\tau = 0.25$, whereas, when we perform interim selection at $\tau = 0.5$, we only need 83 subjects to achieve 80% power. As $\tau$ or $\epsilon$ increases, the required sample size to reach the targeted power decreases.

## 4.6 Normal Approximation and Estimation

### 4.6.1 Normal Approximation

In this section, we graphically demonstrated that the distribution of $W^*$ in Lemma 4.2.1 can be approximated by a mixture of three normal distributions. To reduce redundancy, we only provided detailed work to verify the first component of the distribution of $W^*$. The other two components can be verified in a similar fashion. It has been shown in the previous

Table 4.10: Estimated sample size for $\delta_1 \neq \delta_2 \neq \delta_3$ when $\sigma_Y = 1$ and $\alpha = 0.025$ with targeted power of 80% or 90% by $\tau, \epsilon, \delta_1, \delta_2$ and $\delta_3$ when same endpoints used at the interim and the final analysis.

| | $\tau = 0.25$ | | | | | |
|---|---|---|---|---|---|---|
| | $\epsilon = 0$ | | | $\epsilon = 0.1$ | | |
| $(\delta_1, \delta_2, \delta_3)$ | (0.1,0.2,0.3) | (0.1,0.3,0.5) | (0.1,0.4,0.5) | (0.1,0.2,0.3) | (0.1,0.3,0.5) | (0.1,0.4,0.5) |
| $1-\beta = 0.8$ | 259 | 95 | 83 | 138 | 65 | 57 |
| $1-\beta = 0.9$ | 357 | 130 | 111 | 190 | 90 | 77 |
| | $\tau = 0.5$ | | | | | |
| | $\epsilon = 0$ | | | $\epsilon = 0.1$ | | |
| $(\delta_1, \delta_2, \delta_3)$ | (0.1,0.2,0.3) | (0.1,0.3,0.5) | (0.1,0.4,0.5) | (0.1,0.2,0.3) | (0.1,0.3,0.5) | (0.1,0.4,0.5) |
| $1-\beta = 0.8$ | 228 | 83 | 76 | 124 | 57 | 52 |
| $1-\beta = 0.9$ | 302 | 110 | 100 | 165 | 76 | 69 |
| | $\tau = 0.75$ | | | | | |
| | $\epsilon = 0$ | | | $\epsilon = 0.1$ | | |
| $(\delta_1, \delta_2, \delta_3)$ | (0.1,0.2,0.3) | (0.1,0.3,0.5) | (0.1,0.4,0.5) | (0.1,0.2,0.3) | (0.1,0.3,0.5) | (0.1,0.4,0.5) |
| $1-\beta = 0.8$ | 216 | 79 | 73 | 117 | 54 | 50 |
| $1-\beta = 0.9$ | 283 | 104 | 96 | 154 | 71 | 65 |

section that the first part of the density function of $W^*$ can be written as:

$$f_1(z) = \frac{\sqrt{\frac{3}{4} - \eta^2}}{p_1}\phi(z)\int_{-\infty}^{\frac{\lambda^{(12)}+\eta z}{\sqrt{1-\eta^2}}} \phi(s)\Phi\left(\frac{\left(\lambda^{(13)} + \eta z\right)\sqrt{1-\eta^2} - (\frac{1}{2} - \eta^2)s}{\sqrt{\frac{3}{4} - \eta^2}}\right)ds.$$

where

$$p_1 = Pr\left(\bar{X}_{n_1}^{(1)} = max(\bar{X}_{n_1}^{(1)}, \bar{X}_{n_1}^{(2)}, \bar{X}_{n_1}^{(3)})\right)$$

$$= Pr(\bar{X}_{n_1}^{(1)} > \bar{X}_{n_1}^{(2)}, \bar{X}_{n_1}^{(1)} > \bar{X}_{n_1}^{(3)})$$

$$= \int Pr(\bar{X}_{n_1}^{(1)} > \bar{X}_{n_1}^{(2)}, \bar{X}_{n_1}^{(1)} > \bar{X}_{n_1}^{(3)} | \bar{X}_{n_1}^{(1)} = x_1) f_{\bar{X}_{n_1}^{(1)}}(x_1) dx_1$$

$$
\begin{aligned}
&= \int Pr(x_1 > \bar{X}_{n_1}^{(2)}, x_1 > \bar{X}_{n_1}^{(3)}) f_{\bar{X}_{n_1}^{(1)}}(x_1) dx_1 \\
&= \int Pr(\bar{X}_{n_1}^{(2)} < x_1) Pr(\bar{X}_{n_1}^{(3)} < x_1) f_{\bar{X}_{n_1}^{(1)}}(x_1) dx_1 \\
&= \int F_{\bar{X}_{n_1}^{(2)}}(x_1) F_{\bar{X}_{n_1}^{(3)}}(x_1) f_{\bar{X}_{n_1}^{(1)}}(x_1) dx_1.
\end{aligned}
$$

Hence the mean can be calculated as follows:

$$
\begin{aligned}
\mu_1 &= \int_{-\infty}^{\infty} z f_1(z) dz \\
&= \frac{\sqrt{\frac{3}{4} - \eta^2}}{p_1} \int_{-\infty}^{\infty} z\phi(z) \int_{-\infty}^{\frac{\lambda^{(12)} + \eta z}{\sqrt{1-\eta^2}}} \phi(s) \Phi\left(\frac{\left(\lambda^{(13)} + \eta z\right)\sqrt{1-\eta^2} - (\frac{1}{2} - \eta^2)s}{\sqrt{\frac{3}{4} - \eta^2}}\right) ds dz,
\end{aligned}
$$

and the variance can be calculated as follows:

$$
\begin{aligned}
\sigma_1^2 &= E(z^2) - \mu_1^2 \\
&= \frac{\sqrt{\frac{3}{4} - \eta^2}}{p_1} \int_{-\infty}^{\infty} z^2\phi(z) \int_{-\infty}^{\frac{\lambda^{(12)} + \eta z}{\sqrt{1-\eta^2}}} \phi(s) \Phi\left(\frac{\left(\lambda^{(13)} + \eta z\right)\sqrt{1-\eta^2} - (\frac{1}{2} - \eta^2)s}{\sqrt{\frac{3}{4} - \eta^2}}\right) ds dz - \mu_1^2.
\end{aligned}
$$

By numerical computation of the mean and the variance, we plotted the density function of the normal distribution with mean $\mu_1$ and variance $\sigma_1^2$, and compared it with the curve of $f_1(z)$ with $\eta = \frac{\sqrt{\tau}}{2}\rho = 0.35, \lambda^{(12)} = \sqrt{\frac{n_1}{2\sigma_X^2}}(\nu_1^X - \nu_2^X) = 0, \lambda^{(13)} = \sqrt{\frac{n_1}{2\sigma_X^2}}(\nu_1^X - \nu_3^X) = 0$. Results in Figure 4.19 showed that the two pdf curves are very close, suggesting $f_1(z)$ can be reasonably approximated by the normal distribution with the same mean and variance of $f_1(z)$.

### 4.6.2 Estimation

Similar to Shun et al. (2008), we propose to estimate the final treatment effect $\Delta$, by $\hat{\Delta}_n$ :

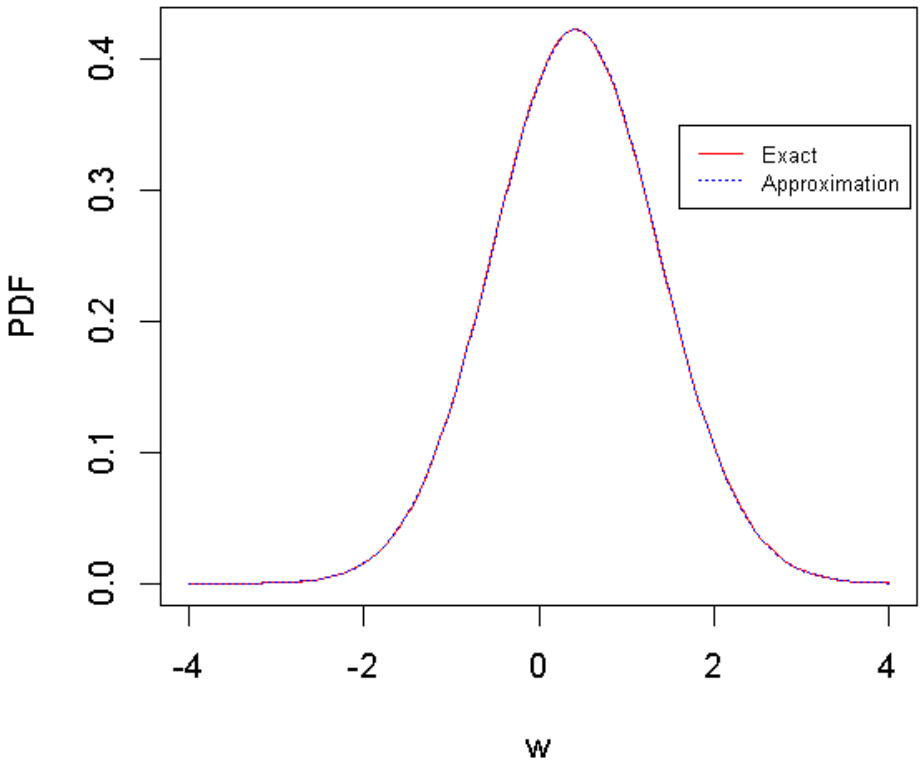
$$
\hat{\Delta}_n = \begin{cases}
\bar{Y}_n^{(1)} - \bar{Y}_n^{(0)}, & \text{if } \bar{X}_{n_1}^{(1)} = max(\bar{X}_{n_1}^{(1)}, \bar{X}_{n_1}^{(2)}, \bar{X}_{n_1}^{(3)}), \\
\bar{Y}_n^{(2)} - \bar{Y}_n^{(0)}, & \text{if } \bar{X}_{n_1}^{(2)} = max(\bar{X}_{n_1}^{(1)}, \bar{X}_{n_1}^{(2)}, \bar{X}_{n_1}^{(3)}), \\
\bar{Y}_n^{(3)} - \bar{Y}_n^{(0)}, & \text{if } \bar{X}_{n_1}^{(3)} = max(\bar{X}_{n_1}^{(1)}, \bar{X}_{n_1}^{(2)}, \bar{X}_{n_1}^{(3)}),
\end{cases} \tag{4.20}
$$

Let $\hat{\Delta}_n^{\delta_j}$ denote the conditional treatment effect when $j^{th}$ treatment is selected at the interim, where $\delta_j$ denotes the $j^{th}$ treatment effect for $j = 1, 2,$ or 3. Using the normal approximation results in Section 4.6.1 and the same arguments of Shun et al. (2008), the $(1-2\alpha)100\%$ "conditional" confidence interval for $\delta_j$ can be approximated by

$$\hat{\Delta}_n^{\delta_j} - \sqrt{\frac{2\sigma_Y^2}{n}}(\mu_j + \sigma_j z_\alpha) < \delta_j < \hat{\Delta}_n^{\delta_j} - \sqrt{\frac{2\sigma_Y^2}{n}}(\mu_j - \sigma_j z_\alpha).$$

Note that $\mu_j$ and $\sigma_j$ can be replaced with the estimates using the interim data as shown in Shun et al. (2008). The confidence intervals can also be derived based on the exact distributions in a similar fashion.

Figure 4.19: Comparison of $f_1(z)$ and the normal pdf with mean $\mu_1 = 0.42$ and $\sigma_1 = 0.94$ when $\nu_1^X - \nu_2^X = 0, \nu_1^X - \nu_3^X = 0, \delta_1 = 0.1, \delta_2 = 0.2, \delta_3 = 0.3, \tau = 0.5, n = 50, \sigma_Y = 1, \sigma_X = 1, \rho = 1. \epsilon = 0.$



# Chapter 5

# Summary, Discussion and Future work

## 5.1 Summary

In Chapter 3 of this dissertation, we developed the methodology for conducting a non-inferiority trial of a two-stage winner design with two experimental treatment and one active control arms. Two scenarios are discussed depending on whether a surrogate endpoint is used at the interim analysis or not. The methodology is summarized as follows:

1. When a surrogate endpoint is used at the interim analysis, the final test statistic differ from the test statistic for superiority hypothesis in Shun et al. (2008) by a constant. We assume the interim and the final endpoints are normally distributed and the variances for the surrogate and the final endpoints are known and equal. The exact distribution for the final test statistic is derived and the critical value that control the size of type I error is studied. Numerical calculation of sample size and power are summarized and shown in tables and graphs.

   Since the final test statistic in Shun et al. can be approximated by normal distributions, the test statistic for non-inferiority hypothesis can also be approximated by normal distributions. Sample size estimations based on the exact distribution and the normal approximation are compared under two scenarios. The first scenario is to pre-specify the information time $\tau$. The second scenario is to pre-specify the winning probability $p$. In both scenarios, the sample size estimates based on the exact distribution and the normal approximation are almost identical. Mathematica programs are provided in Appendix D for solving the sample size estimates using the normal approximation.

2. When a surrogate endpoint is not used at the interim analyses, with the same trial

settings and assumptions, it is shown that the distribution of the final test statistic take the same form regardless when a surrogate endpoint is used or not. However, under this scenario, the parameter $\lambda$ is a function of other parameters, more specifically $\lambda = 2\eta(w_1 - w_2)$. The critical value that control the size of type I error is studied. Numerical calculation of sample size and power are summarized and shown in tables and graphs.

The distribution of the final test statistic can also be approximated by normal distributions under this scenario. Sample size estimations based on the exact distribution and the normal approximation are compared under two scenarios. The first scenario is to pre-specify the information time $\tau$. The second scenario is to pre-specify the winning probability $p$. In both scenarios, the sample size estimates based on the exact distribution and the normal approximation are almost identical. Mathematica programs are provided in Appendix D for solving the sample size estimates using the normal approximation.

In Chapter 4 of this dissertation, we extended the methodology of Chapter 3 to a non-inferiority trial with three experimental treatment and one control arms using a two-stage winner design. The trial setting and assumptions are similar to those in Chapter 3. Superiority hypothesis is considered as a special case of non-inferiority hypothesis when non-inferiority margin $\epsilon = 0$. The results are also discussed depending on whether a surrogate endpoint is used at the interim analysis or not. We summarize the methodology as follows:

1. We first study the case when surrogate endpoints is used in the interim selection. The distribution of the final test statistic is derived for a two-stage winner design with three experimental treatments and one active control. We assume that the interim and the final endpoints are normally distributed with known and equal variances. The tail probability of the test statistic are studied and their related statistical properties are discussed. Numerical calculation of sample size and power are summarized and shown in tables and graphs.

2. When the endpoints for the interim and the final analyses are the same, with the same trial settings and assumptions, the distribution function of the final test statistic is in

the same form as when a surrogate endpoint is used at interim analysis. The difference are in the parameters $\lambda^{(12)}$, and $\lambda^{(13)}$, more specifically, $\lambda^{(12)} = 2\eta(w_1 - w_2), \lambda^{(13)} = 2\eta(w_1 - w_3))$. The tail probability of the test statistic are studied and their related statistical properties are discussed. Numerical calculation of sample size and power are summarized and shown in tables and graphs.

## 5.2 Discussion

In Chapter 3, based on the findings in Shun et al. (2008), we extended the methodology to two-stage winner designs for non-inferiority trials. The motivation of two-stage winner design in non-inferiority trial is from an example of a clinical trial with two experimental treatment versus one active control arms. LUCENTIS® has been approved as an effective treatment for improving visual acuity for patients with age-related macular degeneration (AMD) and it has been on the market for other indications. However, the recommended monthly regimen with invasive eye injection make a big burden for this patient population as most of patients were the elderly. More recently, investigators are seeking alternative treatment regimen that will achieve the same efficacy as the approved one but patient do not need to come in for the monthly injections. Therefore, it is more appropriate to compare the new dosing regimens with the approved dosing regimen (active control) than with the placebo directly. Because one needs only to demonstrate the new dosing regimen is no less effective than the approved regimen, it is preferable to employ the non-inferiority test for this purpose.

Based on this real example, it came to our mind to develop methodology for conducting non-inferiority trials. Because the drug have been approved on the market, we did not consider to design a traditional phase II trials to select one or few promising doses. Recently, there has been considerable interest in adaptive seamless designs that combine the initial "selection" phase followed by a hypothesis testing phase. Adaptive seamless designs offer the flexibility and allows for statistical inference to depend on data from both phases while maintaining and controlling the overall false positive rate. Among the various choices of adaptive seamless design, two-stage seamless adaptive design are frequently used due to it's simplicity. With the appropriate design and good candidate of surrogate endpoints, we can

save a lot of time and resources in clinical development.

With the above consideration in mind, we proposed to conduct a two-stage winner design for non-inferiority trials. The clinical endpoint is assumed to be normally distributed and one interim selection is planned to select a "winner" between the two experimental arms. We assume the interim and final endpoints are normally distributed with known and equal variances. The hypothesis we are interested in is whether one of the high dose of LUCENTIS® with different treatment regimen is not inferior to the approved LUCENTIS® on the market. If we can reject the null hypothesis and prove that one of the experimental treatments is non-inferior to the approved drug (active control), patients will be benefit from receiving a less frequent dosing regimen with invasive injections in the eyes. By using a two-stage winner design, we can save time and resources for conducting a trial.

We first propose a methodology to conduct a non-inferiority trial with two experimental treatment and one active control arms. The methodology is then extended to a trial with three experimental treatment and one active control arms and both superiority and non-inferiority hypotheses are considered. In both methodologies, we discussed two commonly used scenarios depends on whether a surrogate endpoint is used at the interim analysis or not. The first scenario is a surrogate endpoint is used at the interim analysis, in other words, when different endpoints are used at the interim and the final analyses. The second scenario is when a surrogate endpoint is not used, in other words, same endpoints are used at the interim and the final analyses. The distribution of the final test statistics take the same form under these two scenarios, but the underlying definition of the parameters are different.

In a non-inferiority trial with two experimental treatment and one active control arms, when surrogate endpoints is used, the distribution of the final test statistic has a constant shift from the one been studied in Shun et al.(2008). The final test statistic is a mixture of two density functions. Each of the density function can be approximated by a normal distribution, respectively. And the final test statistic can be approximated by a mixture of two normal distributions. We studied the tail probability of the final test statistic $\gamma$ function under $H_0$. In order to determine the critical value that control the type I error rate, we proved two lemmas related to the $\gamma(b; \lambda, w_1, w_2, \eta)$ function when surrogate endpoint is

used: 1) $\gamma(b;\lambda,w_1,w_2,\eta)$ is monotonically increasing with $w_1$ and $w_2$; 2) $\gamma(b;\lambda,w_1,w_2,\eta)$ is maximized at $\lambda=\frac{w_1-w_2}{2\eta}$ when $\rho>0$. When the same endpoints are used at the interim and the final analyses, we found that the $\gamma(b;w_1,w_2,\eta)$ function is maximized at the boundary of $w_1$ and $w_2$. Using these lemmas and theorem, we found the critical value that control the type I error rate in a non-inferiority trial of a two-stage winner design. It turns out that this critical value in a non-inferiority trial is the same as that for a superiority trial and regardless when surrogate endpoint is used or not.

The power and the sample size of the design in association with the parameters $\tau$, $\rho$, $\epsilon$ and $\lambda$ are studied. The power we discussed in this dissertation is the power of rejecting the null hypothesis, not the probability of selecting the right winner. When a surrogate endpoint is used, if the treatment effect for the surrogate endpoint is in the opposite direction as the final endpoint, the power will be smaller. As the non-inferiority margin increases, so is the power. Compared to $\tau$ and $\epsilon$, $\rho$ plays a small role on the power and the sample size. When no surrogate endpoint is used, if the two experimental treatments are close, we will need a larger sample size to achieve the targeted power as compare to the case when the two experimental treatments are far away. When we perform the interim selection at a later time, that chances that we will select the right "winner" is bigger, and therefore, we will need only smaller sample size to achieved the targeted power as compare to the case when we perform interim selection early. Again, as the non-inferiority margin increases, the power increases.

The final test statistics of a two-stage winner design for non-inferiority hypothesis only differ by a constant with the ones for superiority hypothesis. Since the final test statistic for superiority hypothesis in Shun et al.(2008) had been proved can be approximated by normal distributions, the final test statistic for non-inferiority hypothesis can be approximated by normal distributions as well. Sample size estimation using both the exact distribution and the normal approximation are compared and we found that the results from the two methods are almost identical. When we do the sample size estimation using the normal approximation, it is not a trivial process because it involve iteration process to solve non-linear equations. With today's computing power, computer programs to solve the sample size from the exact distribution is much easier as compare to the normal approximation.

The second part of this dissertation extent the work in chapter 3 to the non-inferiority trial with three active treatment arms versus one active control. The trial settings and the assumptions are the same as the ones with two active treatment arms. Both superiority and non-inferiority hypotheses are considered. The distribution of the final test statistic for superiority only differ by a constant with the one for non-inferiority hypothesis. We studied the tail probability $\gamma$ functions. We proved two lemmas related to the $\gamma(b; \lambda^{(12)}, \lambda^{(13)}, w_1, w_2, w_3, \eta)$ function when surrogate endpoint is used: 1) Given $b, \eta, \lambda^{(12)}$ and $\lambda^{(13)}$, $\gamma(b; \lambda^{(12)}, \lambda^{(13)}, w_1, w_2, w_3, \eta)$ is monotonically increasing with $w_1, w_2$ and $w_3$; 2) given $b, \eta, w_1, w_2$ and $w_3$, $\lambda^{(12)} = \lambda^{(13)} = 0$ is a critical point of $\gamma(b; \lambda^{(12)}, \lambda^{(13)}, w_1, w_2, w_3, \eta)$. When the same endpoints are used at the interim and the final analyses, we use numerical justification to show that given $b$ and $\eta$, $\gamma(b; w_1, w_2, w_3, \eta)$ is monotonically increase with respect to $w_1, w_2$ and $w_3$. Using these lemmas and property, we found the critical value that control the type I error rate in a non-inferiority trial of a two-stage winner design. It turns out that this critical value in a non-inferiority trial is the same as that for a superiority trial and regardless when surrogate endpoint is used or not.

The power and the sample size of the design in association with the parameters $\tau$, $\rho$, $\epsilon$, $\lambda^{(12)}$, and $\lambda^{(13)}$ are studied. When a surrogate endpoint is used, if the treatment effect for the surrogate endpoint is in the opposite direction as the final endpoint, the power will be smaller. As the non-inferiority margin increases, so is the power. Compared to $\tau$ and $\epsilon$, $\rho$ plays a small role on the power and the sample size. Similar findings as the trial with two experimental treatment and one active control. When no surrogate endpoint is used, if the two experimental treatments are close, we will need a larger sample size to achieve the targeted power as compare to the case when the two experimental treatments are far away. When we perform the interim selection at a later time, that chances that we will select the right "winner" is bigger, and therefore, we will need only smaller sample size to achieved the targeted power as compare to the case when we perform interim selection early. Again, as the non-inferiority margin increases, the power increases.

The normality assumption for the surrogate or the final endpoints is not critical for the application of the proposed approach, as long as the interim and the final test statistics are normally distributed, the methodology can be applied with given correlation $\eta$ between

the interim and the final test statistics. The endpoints that we discussed in this dissertation mainly focus on the efficacy measurements. However, proper estimation of the safety endpoints should also be considered. In addition, the selection of appropriate surrogate endpoints is not the focus in this dissertation, we only assume the surrogate endpoint and final endpoint are correlated with a correlation. Monotonicity of the tail probability of the test statistic is yet to be proved theoretically. Due to the variety of concepts that used in two-stage winner design and various clinical settings, there are some interesting issues remain unresolved.

## 5.3 Future Work

With the above discussion in mind, the following directions are considered as important potentials in the future.

- Theoretical justification for the monotonicity property of the $\gamma$ functions in Chapter 4.
- Theoretically prove normal approximation of the distribution of $W^*$ in Chapter 4.
- In this dissertation, we only consider the endpoints followed normal distributions. A future work is to study alternatives, such as binomial or survival outcomes.
- Generalize the designs proposed in this thesis to more than three experimental treatments.
- We only select one winner at the interim analysis, a potential research direction is to study designs that select more-than-one winners or drop-the-losers.

# Appendices

# Appendix A

# Useful formulas and properties for the density function of the standard normal distribution

**Lemma A.1.**

$$\begin{aligned}
&\int_{-\infty}^{c} \phi(t)\phi(a+bt)dt \\
&= \int_{-\infty}^{c} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}[\frac{a^2}{(1+b^2)}]} \phi(t\sqrt{1+b^2} + \frac{ab}{\sqrt{1+b^2}})dt, \\
&= \frac{1}{\sqrt{1+b^2}} \phi(\frac{a}{\sqrt{1+b^2}}) \Phi(c\sqrt{1+b^2} + \frac{ab}{\sqrt{1+b^2}}),
\end{aligned}$$

*where $\Phi(.)$ and $\phi(.)$ is the cumulative distribution function (CDF) and probability density function (PDF) of the standard normal distribution.*

*Proof.*

$$\begin{aligned}
&\int_{-\infty}^{c} \phi(t)\phi(a+bt)dt \\
&= \int_{-\infty}^{c} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(a+bt)^2} dt \\
&= \int_{-\infty}^{c} (\frac{1}{\sqrt{2\pi}})^2 e^{-\frac{1}{2}(t^2+a^2+2abt+b^2t^2)} dt \\
&= \int_{-\infty}^{c} (\frac{1}{\sqrt{2\pi}})^2 e^{-\frac{1}{2}[t^2(1+b^2)+2abt+a^2]} dt \\
&= \int_{-\infty}^{c} (\frac{1}{\sqrt{2\pi}})^2 e^{-\frac{1}{2}(\sqrt{1+b^2})^2[t^2+\frac{2abt}{1+b^2}+\frac{a^2}{1+b^2}]} dt \\
&= \int_{-\infty}^{c} (\frac{1}{\sqrt{2\pi}})^2 e^{-\frac{1}{2}(\sqrt{1+b^2})^2[t^2+2t(\frac{ab}{1+b^2})+(\frac{ab}{1+b^2})^2-(\frac{ab}{1+b^2})^2+\frac{a^2}{1+b^2}]} dt \\
&= \int_{-\infty}^{c} (\frac{1}{\sqrt{2\pi}})^2 e^{-\frac{1}{2}(\sqrt{1+b^2})^2[\frac{a^2}{1+b^2}-(\frac{ab}{1+b^2})^2+(t+\frac{ab}{1+b^2})^2]} dt \\
&= \int_{-\infty}^{c} (\frac{1}{\sqrt{2\pi}})^2 e^{-\frac{1}{2}(\sqrt{1+b^2})^2[\frac{a^2}{1+b^2}-(\frac{ab}{1+b^2})^2]} e^{-\frac{1}{2}(\sqrt{1+b^2})^2(t+\frac{ab}{1+b^2})^2} dt
\end{aligned}$$

$$= \int_{-\infty}^{c} (\frac{1}{\sqrt{2\pi}}) e^{-\frac{1}{2}(\sqrt{1+b^2})^2[\frac{a^2}{(1+b^2)^2}]} \phi(t\sqrt{1+b^2} + \frac{ab}{\sqrt{1+b^2}}) dt$$
$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}[\frac{a^2}{(1+b^2)}]} \int_{-\infty}^{c} \phi(t\sqrt{1+b^2} + \frac{ab}{\sqrt{1+b^2}}) dt.$$

Let

$$T = t\sqrt{1+b^2} + \frac{ab}{\sqrt{1+b^2}}, \ \frac{dT}{dt} = \sqrt{1+b^2},$$

Hence,

$$\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}[\frac{a^2}{(1+b^2)}]} \int_{-\infty}^{c} \phi(t\sqrt{1+b^2} + \frac{ab}{\sqrt{1+b^2}}) dt$$
$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}[\frac{a^2}{(1+b^2)}]} \int_{-\infty}^{\frac{ab}{\sqrt{1+b^2}} + c\sqrt{1+b^2}} \phi(T) \frac{dT}{\sqrt{1+b^2}}$$
$$= \frac{1}{\sqrt{2\pi(1+b^2)}} e^{-\frac{1}{2}[\frac{a^2}{(1+b^2)}]} \Phi(c\sqrt{1+b^2} + \frac{ab}{\sqrt{1+b^2}})$$
$$= \frac{1}{\sqrt{1+b^2}} \phi(\frac{a}{\sqrt{1+b^2}}) \Phi(c\sqrt{1+b^2} + \frac{ab}{\sqrt{1+b^2}})$$

□

**Lemma A.2.** *For any $\Delta \geq 0$, let $G_\Delta(x) = \Phi(x+\Delta) + \Phi(x-\Delta) - 2\Phi(x)$, then $G_\Delta(x)$ is an odd function, and for any $x \leq 0$.*

$$G_\Delta(x) = \Phi(x+\Delta) + \Phi(x-\Delta) - 2\Phi(x) \geq 0$$

*where $\Phi(.)$ is the cumulative distribution function of the standard normal distribution.*

*Proof*

$$G_\Delta(-x) = \Phi(-x+\Delta) + \Phi(-x-\Delta) - 2\Phi(-x)$$
$$= \int_{-x}^{-x+\Delta} \phi(z)\, dz + \int_{-x}^{-x-\Delta} \phi(z)\, dz$$
$$= -\int_{x}^{x-\Delta} \phi(z)\, dz - \int_{x}^{x+\Delta} \phi(z)\, dz = -G_\Delta(x),$$

so $G_\Delta(x)$ is an odd function. Furthermore we have

$$\begin{aligned} G_\Delta(x) &= \Phi(x+\Delta) + \Phi(x-\Delta) - 2\Phi(x) \\ &= \int_x^{x-\Delta} \phi(z)\,dz + \int_x^{x+\Delta} \phi(z)\,dz \\ &= \int_x^{x+\Delta} \phi(z)\,dz - \int_{x-\Delta}^{x} \phi(z)\,dz \\ &= \int_x^{x+\Delta} \phi(z)\,dz - \int_x^{x+\Delta} \phi(\Delta - z)\,dz \\ &= \int_x^{x+\Delta} (\phi(z) - \phi(\Delta - z))\,dz \end{aligned}$$

$$\frac{(\Delta - z)^2}{2} = \frac{z^2 - 2z\Delta + \Delta^2}{2} = \frac{z^2}{2} - z\Delta + \frac{\Delta^2}{2} = \frac{z^2}{2} - \Delta(z - \frac{\Delta}{2})$$

$$\phi(z) - \phi(\Delta - z) = \frac{1}{\sqrt{2\pi}} \left( e^{-\frac{z^2}{2}} - e^{-\frac{(\Delta - z)^2}{2}} \right) = \phi(z)\,(1 - e^{\Delta(z - \frac{\Delta}{2})}) \geq 0$$

for $z \leq \frac{\Delta}{2}$. Therefore for $x \leq -\frac{\Delta}{2}$, $x + \Delta \leq \frac{\Delta}{2}$

$$G_\Delta(x) = \int_x^{x+\Delta} (\phi(z) - \phi(\Delta - z))\,dz \geq 0$$

and for $-\frac{\Delta}{2} \leq x \leq 0$, we have $-x \leq \frac{\Delta}{2} \leq x + \Delta \leq \Delta$, since

$$\int_{-x}^{x+\Delta} (\phi(z) - \phi(\Delta - z))\,dz = \int_{-x}^{x+\Delta} \phi(z)\,dz + \int_{x+\Delta}^{-x} \phi(y)\,dy = 0,$$

by letting $y = \Delta - z$ in the above second term, and

$$\begin{aligned} G_\Delta(x) &= \int_x^{x+\Delta} (\phi(z) - \phi(\Delta - z))\,dz \\ &= \int_x^{-x} (\phi(z) - \phi(\Delta - z))\,dz + \int_{-x}^{x+\Delta} (\phi(z) - \phi(\Delta - z))\,dz \\ &= \int_x^{-x} (\phi(z) - \phi(\Delta - z))\,dz \geq 0. \end{aligned}$$

Therefore, for any $x \leq 0$, $G_\Delta(x) \geq 0$. □

# Appendix B

# Proofs in Chapter 3

## B.1 Proof of Lemma 3.3.1

*Proof.* For any $w_1$ and $w_2$,

$$\frac{\partial}{\partial w_1}\gamma(b;\lambda,w_1,w_2,\eta)=\Phi\left(\frac{\lambda+\eta\left(b-w_1-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon\right)}{\sqrt{1-\eta^2}}\right)\phi\left(b-w_1-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon\right)>0,$$

and

$$\frac{\partial}{\partial w_2}\gamma(b;\lambda,w_1,w_2,\eta)=\Phi\left(\frac{-\lambda+\eta\left(b-w_2-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon\right)}{\sqrt{1-\eta^2}}\right)\phi\left(b-w_2-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon\right)>0.$$

Therefore, given $\lambda$ and $b$, $\gamma(b;\lambda,w_1,w_2,\eta)$ is monotonically increasing with $w_1$ and $w_2$.

□

## B.2 Proof of Lemma 3.3.2

*Proof.*

$$\begin{aligned}\frac{\partial}{\partial\lambda}\gamma(b;\lambda,w_1,w_2,\eta)=&\int_{-\infty}^{b-w_1-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon}\frac{1}{\sqrt{1-\eta^2}}\phi\left(\frac{\lambda+\eta z}{\sqrt{1-\eta^2}}\right)\phi(z)\,dz\\&-\int_{-\infty}^{b-w_2-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon}\frac{1}{\sqrt{1-\eta^2}}\phi\left(\frac{-\lambda+\eta z}{\sqrt{1-\eta^2}}\right)\phi(z)\,dz,\end{aligned}$$

let

$$a=\frac{\lambda}{\sqrt{1-\eta^2}},b=\frac{\eta}{\sqrt{1-\eta^2}},$$

$$\sqrt{1+b^2} = \sqrt{1+\frac{\eta^2}{1-\eta^2}} = \sqrt{\frac{1-\eta^2+\eta^2}{1-\eta^2}} = \frac{1}{\sqrt{1-\eta^2}},$$
$$\frac{a}{\sqrt{1+b^2}} = \frac{\lambda}{\sqrt{1-\eta^2}}\sqrt{1-\eta^2} = \lambda$$
$$\frac{ab}{\sqrt{1+b^2}} = \frac{\lambda}{\sqrt{1-\eta^2}}\frac{\eta}{\sqrt{1-\eta^2}}\sqrt{1-\eta^2} = \frac{\lambda\eta}{\sqrt{1-\eta^2}},$$

by Lemma A.1, we have

$$\int_{-\infty}^{b-w_1-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon} \frac{1}{\sqrt{1-\eta^2}}\phi\left(\frac{\lambda+\eta z}{\sqrt{1-\eta^2}}\right)\phi(z)\,dz$$
$$= \frac{1}{\sqrt{1-\eta^2}}\sqrt{1-\eta^2}\phi(\lambda)\,\Phi\left[\frac{\left(b-w_1-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon\right)+\lambda\eta}{\sqrt{1-\eta^2}}\right].$$

Similarly, let

$$a = \frac{-\lambda}{\sqrt{1-\eta^2}}, b = \frac{\eta}{\sqrt{1-\eta^2}},$$
$$\sqrt{1+b^2} = \sqrt{1+\frac{\eta^2}{1-\eta^2}} = \sqrt{\frac{1-\eta^2+\eta^2}{1-\eta^2}} = \frac{1}{\sqrt{1-\eta^2}},$$
$$\frac{a}{\sqrt{1+b^2}} = \frac{\lambda}{\sqrt{1-\eta^2}}\sqrt{1-\eta^2} = -\lambda$$
$$\frac{ab}{\sqrt{1+b^2}} = \frac{-\lambda}{\sqrt{1-\eta^2}}\frac{\eta}{\sqrt{1-\eta^2}}\sqrt{1-\eta^2} = \frac{-\lambda\eta}{\sqrt{1-\eta^2}},$$

by Lemma A.1, we have

$$\int_{-\infty}^{b-w_2-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon} \frac{1}{\sqrt{1-\eta^2}}\phi\left(\frac{-\lambda+\eta z}{\sqrt{1-\eta^2}}\right)\phi(z)\,dz$$
$$= \frac{1}{\sqrt{1-\eta^2}}\sqrt{1-\eta^2}\phi(-\lambda)\,\Phi\left[\frac{\left(b-w_2-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon\right)-\lambda\eta}{\sqrt{1-\eta^2}}\right].$$

Therefore,

$$\int_{-\infty}^{b-w_1-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon} \frac{1}{\sqrt{1-\eta^2}}\phi\left(\frac{\lambda+\eta z}{\sqrt{1-\eta^2}}\right)\phi(z)\,dz$$

$$
\begin{aligned}
&- \int_{-\infty}^{b-w_2-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon} \frac{1}{\sqrt{1-\eta^2}} \phi\left(\frac{-\lambda+\eta z}{\sqrt{1-\eta^2}}\right) \phi(z)\, dz \\
&= \phi(\lambda)\, \Phi\left[\frac{\left(b-w_1-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon\right)}{\sqrt{1-\eta^2}} + \frac{\lambda\eta}{\sqrt{1-\eta^2}}\right] \\
&\quad - \phi(-\lambda)\, \Phi\left[\frac{\left(b-w_2-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon\right)}{\sqrt{1-\eta^2}} - \frac{\lambda\eta}{\sqrt{1-\eta^2}}\right] \\
&= \phi(\lambda) \left\{ \Phi\left[\frac{\left(b-w_1-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon\right)+\lambda\eta}{\sqrt{1-\eta^2}}\right] - \Phi\left[\frac{\left(b-w_2-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon\right)-\lambda\eta}{\sqrt{1-\eta^2}}\right] \right\}.
\end{aligned}
$$

When $\rho > 0$, $\eta = \frac{\sqrt{\tau}}{2}\rho > 0$, we have

$$\left(b-w_1-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon\right)+\lambda\eta > \left(b-w_2-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon\right)-\lambda\eta,$$

$$2\lambda\eta > w_1 - w_2,$$

$$\lambda > \frac{w_1-w_2}{2\eta}.$$

Therefore, for any $\lambda > \frac{w_1-w_2}{2\eta}$, we have

$$\Phi\left[\frac{\left(b-w_1-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon\right)+\lambda\eta}{\sqrt{1-\eta^2}}\right] - \Phi\left[\frac{\left(b-w_2-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon\right)-\lambda\eta}{\sqrt{1-\eta^2}}\right] > 0,$$

and for any $\lambda < \frac{w_1-w_2}{2\eta}$, we have

$$\Phi\left[\frac{\left(b-w_1-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon\right)+\lambda\eta}{\sqrt{1-\eta^2}}\right] - \Phi\left[\frac{\left(b-w_2-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon\right)-\lambda\eta}{\sqrt{1-\eta^2}}\right] < 0.$$

This implies that $\gamma(b;\lambda,w_1,w_2,\eta)$ is monotonically increasing with $\lambda$ when $\lambda > \frac{w_1-w_2}{2\eta}$, and monotonically decreasing with $\lambda$ when $\lambda < \frac{w_1-w_2}{2\eta}$, hence the maximum of $\gamma(b;\lambda,w_1,w_2,\eta)$ occurs at $\lambda = \frac{w_1-w_2}{2\eta}$, for any given $w_1$, $w_2$ and $b$ when $\rho > 0$.

□

## B.3 Proof of Lemma 3.3.3

*Proof.* For any $0 \leq \eta < 1$ and $0 \leq u_1 \leq u_2$, let $\Delta = \frac{2\eta(u_2-u_1)}{\sqrt{1-\eta^2}}$, then $\Delta \geq 0$ and by Lemma A.2, $G_\Delta\left(\frac{\eta z}{\sqrt{1-\eta^2}}\right)$ is an odd function of $z$ and $G_\Delta\left(\frac{\eta z}{\sqrt{1-\eta^2}}\right) \geq 0$ for any $z \leq 0$. Therefore

$$
\begin{aligned}
&\int_{-\infty}^{u_1} \Phi\left(\frac{2\eta(u_2-u_1)+\eta z}{\sqrt{1-\eta^2}}\right)\phi(z)\,dz + \int_{-\infty}^{u_2} \Phi\left(\frac{2\eta(u_1-u_2)+\eta z}{\sqrt{1-\eta^2}}\right)\phi(z)\,dz \\
&\quad - 2\int_{-\infty}^{u_1} \Phi\left(\frac{\eta z}{\sqrt{1-\eta^2}}\right)\phi(z)\,dz \\
&= \int_{u_1}^{u_2} \Phi\left(\frac{\eta z}{\sqrt{1-\eta^2}} - \Delta\right)\phi(z)\,dz \\
&\quad + \int_{-\infty}^{u_1} \left(\Phi\left(\frac{\eta z}{\sqrt{1-\eta^2}} + \Delta\right) + \Phi\left(\frac{\eta z}{\sqrt{1-\eta^2}} - \Delta\right) - 2\Phi\left(\frac{\eta z}{\sqrt{1-\eta^2}}\right)\right)\phi(z)\,dz \\
&\geq \int_{-\infty}^{u_1} \left(\Phi\left(\frac{\eta z}{\sqrt{1-\eta^2}} + \Delta\right) + \Phi\left(\frac{\eta z}{\sqrt{1-\eta^2}} - \Delta\right) - 2\Phi\left(\frac{\eta z}{\sqrt{1-\eta^2}}\right)\right)\phi(z)\,dz \\
&= \int_{-\infty}^{u_1} G_\Delta\left(\frac{\eta z}{\sqrt{1-\eta^2}}\right)\phi(z)\,dz \\
&= \int_{-\infty}^{-u_1} G_\Delta\left(\frac{\eta z}{\sqrt{1-\eta^2}}\right)\phi(z)\,dz \\
&\geq 0.
\end{aligned}
$$

□

# Appendix C

# Proofs in Chapter 4

## C.1 Proof of Lemma 4.2.1

*Proof.* From assumptions, $(Z_n^{(1)}, V_{n_1}^{(12)}, V_{n_1}^{(13)})$ follows the multivariate normal distribution with mean $\mu = (w_1, \lambda^{(12)}, \lambda^{(13)})^T$ and variance

$$\Sigma = \begin{pmatrix} 1 & \eta & \eta \\ \eta & 1 & 1/2 \\ \eta & 1/2 & 1 \end{pmatrix}.$$

Therefore, $(V_{n_1}^{(12)}, V_{n_1}^{(13)})^T | Z_n^{(1)}$ follows the bivariate normal distribution with the mean

$$\begin{pmatrix} \lambda^{(12)} \\ \lambda^{(13)} \end{pmatrix} + \begin{pmatrix} \eta \\ \eta \end{pmatrix} \left( Z_n^{(1)} - w_1 \right) = \begin{pmatrix} \lambda^{(12)} + \eta \left( Z^{(1)} - w_1 \right) \\ \lambda^{(13)} + \eta \left( Z^{(1)} - w_1 \right) \end{pmatrix},$$

and the variance

$$\begin{pmatrix} 1 & 1/2 \\ 1/2 & 1 \end{pmatrix} - \begin{pmatrix} \eta \\ \eta \end{pmatrix} (\eta, \eta) = \begin{pmatrix} 1 - \eta^2 & 1/2 - \eta^2 \\ 1/2 - \eta^2 & 1 - \eta^2. \end{pmatrix},$$

i.e.,

$$\begin{pmatrix} V_{n_1}^{(12)} \\ V_{n_1}^{(13)} \end{pmatrix} | Z_n^{(1)} \sim N \left( \begin{pmatrix} \lambda^{(12)} + \eta \left( Z_n^{(1)} - w_1 \right) \\ \lambda^{(13)} + \eta \left( Z_n^{(1)} - w_1 \right) \end{pmatrix}, \begin{pmatrix} 1 - \eta^2 & 1/2 - \eta^2 \\ 1/2 - \eta^2 & 1 - \eta^2 \end{pmatrix} \right).$$

By symmetry, we have

$$\begin{pmatrix} V_{n_1}^{(21)} \\ V_{n_1}^{(23)} \end{pmatrix} |Z_n^{(2)} \sim N\left(\begin{pmatrix} \lambda^{(21)} + \eta\left(Z_n^{(2)} - w_2\right) \\ \lambda^{(23)} + \eta\left(Z_n^{(2)} - w_2\right) \end{pmatrix}, \begin{pmatrix} 1-\eta^2 & 1/2-\eta^2 \\ 1/2-\eta^2 & 1-\eta^2 \end{pmatrix}\right),$$

and

$$\begin{pmatrix} V_{n_1}^{(31)} \\ V_{n_1}^{(32)} \end{pmatrix} |Z^{(3)} \sim N\left(\begin{pmatrix} \lambda^{(31)} + \eta\left(Z_n^{(3)} - w_3\right) \\ \lambda^{(32)} + \eta\left(Z_n^{(3)} - w_3\right) \end{pmatrix}, \begin{pmatrix} 1-\eta^2 & 1/2-\eta^2 \\ 1/2-\eta^2 & 1-\eta^2 \end{pmatrix}\right),$$

where $V_{n_1}^{(ij)} = \sqrt{\frac{n}{2\sigma_X^2}}(\overline{X}_{n_1}^{(i)} - \overline{X}_{n_1}^{(j)})$ and $\lambda^{(ij)} = \sqrt{\frac{n}{2\sigma_X^2}}(\nu_i^X - \nu_j^X)$ for $i \neq j$.

In terms of $V_{n_1}^{(12)}, V_{n_1}^{(13)}, \lambda^{(12)}$ and $\lambda^{(13)}$, we have

$$\begin{aligned}
\begin{pmatrix} V_{n_1}^{(12)} \\ V_{n_1}^{(13)} \end{pmatrix} |Z_n^{(1)} &\sim N\left(\begin{pmatrix} \lambda^{(12)} + \eta\left(Z_n^{(1)} - w_1\right) \\ \lambda^{(13)} + \eta\left(Z_n^{(1)} - w_1\right) \end{pmatrix}, \begin{pmatrix} 1-\eta^2 & 1/2-\eta^2 \\ 1/2-\eta^2 & 1-\eta^2 \end{pmatrix}\right) \\
&= \begin{pmatrix} \lambda^{(12)} + \eta\left(Z_n^{(1)} - w_1\right) \\ \lambda^{(13)} + \eta\left(Z_n^{(1)} - w_1\right) \end{pmatrix} - N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1-\eta^2 & 1/2-\eta^2 \\ 1/2-\eta^2 & 1-\eta^2 \end{pmatrix}\right) \\
&= \begin{pmatrix} \lambda^{(12)} + \eta\left(Z_n^{(1)} - w_1\right) \\ \lambda^{(13)} + \eta\left(Z_n^{(1)} - w_1\right) \end{pmatrix} - \begin{pmatrix} S \\ T \end{pmatrix},
\end{aligned}$$

$$\begin{aligned}
\begin{pmatrix} V_{n_1}^{(12)} \\ V_{n_1}^{(12)} - V_{n_1}^{(13)} \end{pmatrix} |Z_n^{(2)} &= -\begin{pmatrix} V_{n_1}^{(21)} \\ V_{n_1}^{(23)} \end{pmatrix} |Z_n^{(2)} \\
&\sim N\left(\begin{pmatrix} \lambda^{(12)} - \eta\left(Z_n^{(2)} - w_2\right) \\ \lambda^{(12)} - \lambda^{(13)} - \eta\left(Z_n^{(2)} - w_2\right) \end{pmatrix}, \begin{pmatrix} 1-\eta^2 & 1/2-\eta^2 \\ 1/2-\eta^2 & 1-\eta^2 \end{pmatrix}\right) \\
&= \begin{pmatrix} \lambda^{(12)} - \eta\left(Z_n^{(2)} - w_2\right) \\ \lambda^{(12)} - \lambda^{(13)} - \eta\left(Z_n^{(2)} - w_2\right) \end{pmatrix} \\
&\qquad + N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1-\eta^2 & 1/2-\eta^2 \\ 1/2-\eta^2 & 1-\eta^2 \end{pmatrix}\right)
\end{aligned}$$

$$= \begin{pmatrix} \lambda^{(12)} - \eta\left(Z_n^{(2)} - w_2\right) \\ \lambda^{(12)} - \lambda^{(13)} - \eta\left(Z_n^{(2)} - w_2\right) \end{pmatrix} + \begin{pmatrix} S \\ T \end{pmatrix},$$

and

$$\begin{aligned}
\begin{pmatrix} V_{n_1}^{(13)} \\ V_{n_1}^{(13)} - V_{n_1}^{(12)} \end{pmatrix} |Z^{(3)} &= - \begin{pmatrix} V_{n_1}^{(31)} \\ V_{n_1}^{(32)} \end{pmatrix} |Z^{(3)} \\
&\sim N\left( \begin{pmatrix} \lambda^{(13)} - \eta\left(Z_n^{(3)} - w_3\right) \\ \lambda^{(13)} - \lambda^{(12)} - \eta\left(Z_n^{(3)} - w_3\right) \end{pmatrix}, \begin{pmatrix} 1-\eta^2 & 1/2-\eta^2 \\ 1/2-\eta^2 & 1-\eta^2 \end{pmatrix} \right) \\
&= \begin{pmatrix} \lambda^{(13)} - \eta\left(Z_n^{(3)} - w_3\right) \\ \lambda^{(13)} - \lambda^{(12)} - \eta\left(Z_n^{(3)} - w_3\right) \end{pmatrix} \\
&\qquad + N\left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1-\eta^2 & 1/2-\eta^2 \\ 1/2-\eta^2 & 1-\eta^2 \end{pmatrix} \right) \\
&= \begin{pmatrix} \lambda^{(13)} - \eta\left(Z_n^{(3)} - w_3\right) \\ \lambda^{(13)} - \lambda^{(12)} - \eta\left(Z_n^{(3)} - w_3\right) \end{pmatrix} + \begin{pmatrix} S \\ T \end{pmatrix},
\end{aligned}$$

where

$$\begin{pmatrix} S \\ T \end{pmatrix} \sim N\left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1-\eta^2 & 1/2-\eta^2 \\ 1/2-\eta^2 & 1-\eta^2 \end{pmatrix} \right). \tag{C.1.1}$$

Notice that $\begin{pmatrix} S \\ T \end{pmatrix}$ only depends on $\eta$.

From the above results, we can derive the distribution of the test statistic $W$ as follows:

$$\begin{aligned}
F_W(w) &= Pr(W < w) \\
&= Pr\left(Z_n^{(1)} < w, \bar{X}_{n_1}^{(1)} = max(\bar{X}_{n_1}^{(1)}, \bar{X}_{n_1}^{(2)}, \bar{X}_{n_1}^{(3)})\right) \\
&\quad + Pr\left(Z_n^{(2)} < w, \bar{X}_{n_1}^{(2)} = max(\bar{X}_{n_1}^{(1)}, \bar{X}_{n_1}^{(2)}, \bar{X}_{n_1}^{(3)})\right) \\
&\quad + Pr\left(Z_n^{(3)} < w, \bar{X}_{n_1}^{(3)} = max(\bar{X}_{n_1}^{(1)}, \bar{X}_{n_1}^{(2)}, \bar{X}_{n_1}^{(3)})\right) \\
&= \int_{-\infty}^{w} \Big[\phi(z - w_1) P(V_{n_1}^{(12)} > 0, V_{n_1}^{(13)} > 0) \; + \phi(z - w_2) P(V_{n_1}^{(12)} < 0, V_{n_1}^{(13)} > V_{n_1}^{(12)})
\end{aligned}$$

$$
\begin{aligned}
&+\phi(z-w_3)P(V_{n_1}^{(13)}<0, V_{n_1}^{(13)}<V_{n_1}^{(12)})\Big]\,dz \\
=&\int_{-\infty}^{w}\Big[\phi(z-w_1)P(S<\lambda^{(12)}+\eta\,(z-w_1)\,, T<\lambda^{(13)}+\eta\,(z-w_1)) \\
&+\phi(z-w_2)P(S<-\lambda^{(12)}+\eta\,(z-w_2)\,, T<\lambda^{(13)}-\lambda^{(12)}+\eta\,(z-w_2)) \\
&+\phi(z-w_3)P(S<-\lambda^{(13)}+\eta\,(z-w_3)\,, T<\lambda^{(12)}-\lambda^{(13)}+\eta\,(z-w_3))\Big]\,dz \\
=&\int_{-\infty}^{w-w_1}\phi(z)P(S<\lambda^{(12)}+\eta z, T<\lambda^{(13)}+\eta z)dz \\
&+\int_{-\infty}^{w-w_2}\phi(z)P(S<-\lambda^{(12)}+\eta z, T<\lambda^{(13)}-\lambda^{(12)}+\eta z)dz \\
&+\int_{-\infty}^{w-w_3}\phi(z)P(S<-\lambda^{(13)}+\eta z, T<\lambda^{(12)}-\lambda^{(13)}+\eta z)dz.
\end{aligned}
$$

From (C.1.1), we have

$$
S\sim N(0,1-\eta^2), T|S\sim N\left(s\rho, \left(1-\eta^2\right)\left(1-\rho^2\right)\right).
$$

where $\rho=\frac{\frac{1}{2}-\eta^2}{1-\eta^2}, 1-\rho^2=1-\frac{\left(\frac{1}{2}-\eta^2\right)^2}{\left(1-\eta^2\right)^2}=\frac{1-2\eta^2+\eta^4-\frac{1}{4}+\eta^2-\eta^4}{\left(1-\eta^2\right)^2}=\frac{\frac{3}{4}-\eta^2}{\left(1-\eta^2\right)^2}$. Therefore, for any a and b,

$$
\begin{aligned}
P(S<a, T<b)&=\int_{-\infty}^{a}\int_{-\infty}^{b}f(s,t)dtds=\int_{-\infty}^{a}f(s)\int_{-\infty}^{b}f(t|s)dtds \\
&=\int_{-\infty}^{a}\phi(\frac{s}{\sqrt{1-\eta^2}})\int_{-\infty}^{b}\phi(\frac{t-s\rho}{\sqrt{\left(1-\eta^2\right)\left(1-\rho^2\right)}})dtds \\
&=\sqrt{\left(1-\eta^2\right)\left(1-\rho^2\right)}\int_{-\infty}^{a}\phi(\frac{s}{\sqrt{1-\eta^2}})\Phi(\frac{b-s\rho}{\sqrt{\left(1-\eta^2\right)\left(1-\rho^2\right)}})ds \\
&=\sqrt{\frac{\frac{3}{4}-\eta^2}{1-\eta^2}}\int_{-\infty}^{a}\phi(\frac{s}{\sqrt{1-\eta^2}})\Phi(\frac{b-s\rho}{\sqrt{\frac{\frac{3}{4}-\eta^2}{1-\eta^2}}})ds \\
&=\sqrt{(\frac{3}{4}-\eta^2)}\int_{-\infty}^{\frac{a}{\sqrt{1-\eta^2}}}\phi(s')\Phi(\frac{b-s'\sqrt{1-\eta^2}\left(\frac{\frac{1}{2}-\eta^2}{1-\eta^2}\right)}{\sqrt{\frac{\frac{3}{4}-\eta^2}{1-\eta^2}}})ds' \\
&=\sqrt{(\frac{3}{4}-\eta^2)}\int_{-\infty}^{\frac{a}{\sqrt{1-\eta^2}}}\phi(s)\Phi(\frac{b\sqrt{1-\eta^2}-(\frac{1}{2}-\eta^2)s}{\sqrt{\frac{3}{4}-\eta^2}})ds. \qquad \text{(C.1.2)}
\end{aligned}
$$

By (C.1.2), the distribution of $W$ can be further derived as follows:

$$
\begin{aligned}
&\int_{-\infty}^{w-w_1} \phi(z)P(S<\lambda^{(12)}+\eta z, T<\lambda^{(13)}+\eta z)dz \\
&\quad+\int_{-\infty}^{w-w_2} \phi(z)P(S<-\lambda^{(12)}+\eta z, T<\lambda^{(13)}-\lambda^{(12)}+\eta z)dz \\
&\quad+\int_{-\infty}^{w-w_3} \phi(z)P(S<-\lambda^{(13)}+\eta z, T<\lambda^{(12)}-\lambda^{(13)}+\eta z)dz \\
&=\sqrt{\frac{3}{4}-\eta^2}\int_{-\infty}^{w-w_1}\phi(z)\int_{-\infty}^{\frac{\lambda^{(12)}+\eta z}{\sqrt{1-\eta^2}}}\phi(s)\Phi(\frac{\left(\lambda^{(13)}+\eta z\right)\sqrt{1-\eta^2}-(\frac{1}{2}-\eta^2)s}{\sqrt{\frac{3}{4}-\eta^2}})dsdz \\
&\quad+\sqrt{\frac{3}{4}-\eta^2}\int_{-\infty}^{w-w_2}\phi(z)\int_{-\infty}^{\frac{-\lambda^{(12)}+\eta z}{\sqrt{1-\eta^2}}}\phi(s)\Phi(\frac{\left(\lambda^{(13)}-\lambda^{(12)}+\eta z\right)\sqrt{1-\eta^2}-(\frac{1}{2}-\eta^2)s}{\sqrt{\frac{3}{4}-\eta^2}})dsdz \\
&\quad+\sqrt{\frac{3}{4}-\eta^2}\int_{-\infty}^{w-w_3}\phi(z)\int_{-\infty}^{\frac{-\lambda^{(13)}+\eta z}{\sqrt{1-\eta^2}}}\phi(s)\Phi(\frac{\left(\lambda^{(12)}-\lambda^{(13)}+\eta z\right)\sqrt{1-\eta^2}-(\frac{1}{2}-\eta^2)s}{\sqrt{\frac{3}{4}-\eta^2}})dsdz.
\end{aligned}
$$

□

## C.2 Proof of Lemma 4.3.1

*Proof.*

$$
\begin{aligned}
&\frac{\partial}{\partial w_1}\gamma(b;\lambda^{(12)},\lambda^{(13)},w_1,w_2,w_3) \\
&=\sqrt{(\frac{3}{4}-\eta^2)}\int_{-\infty}^{\frac{\lambda^{(12)}+\eta(b-w_1-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon)}{\sqrt{1-\eta^2}}}\phi(b-w_1-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon)\phi(s) \\
&\qquad\times\Phi(\frac{\left[\lambda^{(13)}+\eta(b-w_1-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon)\right]\sqrt{1-\eta^2}-(\frac{1}{2}-\eta^2)s}{\sqrt{\frac{3}{4}-\eta^2}})ds>0.
\end{aligned}
$$

$$
\begin{aligned}
&\frac{\partial}{\partial w_2}\gamma(b;\lambda^{(12)},\lambda^{(13)},w_1,w_2,w_3) \\
&=\sqrt{(\frac{3}{4}-\eta^2)}\int_{-\infty}^{\frac{-\lambda^{(12)}+\eta(b-w_2-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon)}{\sqrt{1-\eta^2}}}\phi(b-w_2-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon)\phi(s)
\end{aligned}
$$

$$\times \Phi(\frac{\left[\lambda^{(13)} - \lambda^{(12)} + \eta(b - w_2 - \sqrt{\frac{n}{2\sigma_Y^2}}\epsilon)\right]\sqrt{1-\eta^2} - (\frac{1}{2} - \eta^2)s}{\sqrt{\frac{3}{4} - \eta^2}})ds > 0.$$

$$\frac{\partial}{\partial w_3}\gamma(b; \lambda^{(12)}, \lambda^{(13)}, w_1, w_2, w_3)$$
$$= \sqrt{(\frac{3}{4} - \eta^2)} \int_{-\infty}^{\frac{-\lambda^{(13)} + \eta(b - w_3 - \sqrt{\frac{n}{2\sigma_Y^2}}\epsilon)}{\sqrt{1-\eta^2}}} \phi(b - w_3 - \sqrt{\frac{n}{2\sigma_Y^2}}\epsilon)\phi(s)$$
$$\times \Phi(\frac{\left[\lambda^{(12)} - \lambda^{(13)} + \eta(b - w_3 - \sqrt{\frac{n}{2\sigma_Y^2}}\epsilon)\right]\sqrt{1-\eta^2} - (\frac{1}{2} - \eta^2)s}{\sqrt{\frac{3}{4} - \eta^2}})ds > 0.$$

Given $b, \lambda^{(12)}$ and $\lambda^{(13)}$, we know that $\gamma(b; \lambda^{(12)}, \lambda^{(13)}, w_1, w_2, w_3)$ is monotonically increasing with $w_1, w_2$ and $w_3$. □

## C.3 Proof of Lemma 4.3.2

*Proof.* By Lemma 4.3.1, given $b, \lambda^{(12)}$ and $\lambda^{(13)}$, we know $\gamma(b; \lambda^{(12)}, \lambda^{(13)}, w_1, w_2, w_3)$ is monotonically increasing with $w_1, w_2$ and $w_3$.

Let $w_1 = w_2 = w_3 = 0$ in $\gamma(b; \lambda^{(12)}, \lambda^{(13)}, w_1, w_2, w_3)$ , we have

$$\gamma(b; \lambda^{(12)}, \lambda^{(13)}, w_1 = w_2 = w_3 = 0)$$
$$= 1 - \sqrt{(\frac{3}{4} - \eta^2)} \int_{-\infty}^{b - \sqrt{\frac{n}{2\sigma_Y^2}}\epsilon} \phi(z) \int_{-\infty}^{\frac{\lambda^{(12)} + \eta z}{\sqrt{1-\eta^2}}} \phi(s)\Phi(\frac{\left(\lambda^{(13)} + \eta z\right)\sqrt{1-\eta^2} - (\frac{1}{2} - \eta^2)s}{\sqrt{\frac{3}{4} - \eta^2}})dsdz$$
$$- \sqrt{(\frac{3}{4} - \eta^2)} \int_{-\infty}^{b - \sqrt{\frac{n}{2\sigma_Y^2}}\epsilon} \phi(z) \int_{-\infty}^{\frac{-\lambda^{(12)} + \eta z}{\sqrt{1-\eta^2}}} \phi(s)\Phi(\frac{\left(\lambda^{(13)} - \lambda^{(12)} + \eta z\right)\sqrt{1-\eta^2} - (\frac{1}{2} - \eta^2)s}{\sqrt{\frac{3}{4} - \eta^2}})dsdz$$
$$- \sqrt{(\frac{3}{4} - \eta^2)} \int_{-\infty}^{b - \sqrt{\frac{n}{2\sigma_Y^2}}\epsilon} \phi(z) \int_{-\infty}^{\frac{-\lambda^{(13)} + \eta z}{\sqrt{1-\eta^2}}} \phi(s)\Phi(\frac{\left(\lambda^{(12)} - \lambda^{(13)} + \eta z\right)\sqrt{1-\eta^2} - (\frac{1}{2} - \eta^2)s}{\sqrt{\frac{3}{4} - \eta^2}})dsdz,$$

by taking the partial derivative with respect to $\lambda^{(12)}$ in $\gamma$ when $w_1 = w_2 = w_3 = 0$, we have

$$\frac{\partial}{\partial \lambda^{(12)}}\gamma(b; \lambda^{(12)}, \lambda^{(13)}, w_1 = w_2 = w_3 = 0)$$

$$= -\sqrt{\frac{\frac{3}{4}-\eta^2}{1-\eta^2}} \int_{-\infty}^{b-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon} \phi(z)\phi(\frac{\lambda^{(12)}+\eta z}{\sqrt{1-\eta^2}})\Phi\left(\frac{\left(\lambda^{(13)}+\eta z\right)\sqrt{1-\eta^2}-(\frac{1}{2}-\eta^2)\left(\frac{\lambda^{(12)}+\eta z}{\sqrt{1-\eta^2}}\right)}{\sqrt{\frac{3}{4}-\eta^2}}\right)dz$$

$$+\sqrt{\frac{\frac{3}{4}-\eta^2}{1-\eta^2}} \int_{-\infty}^{b-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon} \phi(z)\phi(\frac{-\lambda^{(12)}+\eta z}{\sqrt{1-\eta^2}})$$

$$\times\Phi\left(\frac{\left(\lambda^{(13)}-\lambda^{(12)}+\eta z\right)\sqrt{1-\eta^2}-(\frac{1}{2}-\eta^2)\left(\frac{-\lambda^{(12)}+\eta z}{\sqrt{1-\eta^2}}\right)}{\sqrt{\frac{3}{4}-\eta^2}}\right)dz$$

$$+\sqrt{1-\eta^2}\int_{-\infty}^{b-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon}\phi(z)\int_{-\infty}^{\frac{-\lambda^{(12)}+\eta z}{\sqrt{1-\eta^2}}}\phi(s)\phi\left(\frac{\left(\lambda^{(13)}-\lambda^{(12)}+\eta z\right)\sqrt{1-\eta^2}-(\frac{1}{2}-\eta^2)s}{\sqrt{\frac{3}{4}-\eta^2}}\right)dsdz$$

$$-\sqrt{1-\eta^2}\int_{-\infty}^{b-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon}\phi(z)\int_{-\infty}^{\frac{-\lambda^{(13)}+\eta z}{\sqrt{1-\eta^2}}}\phi(s)\phi\left(\frac{\left(\lambda^{(12)}-\lambda^{(13)}+\eta z\right)\sqrt{1-\eta^2}-(\frac{1}{2}-\eta^2)s}{\sqrt{\frac{3}{4}-\eta^2}}\right)dsdz.$$

When $\lambda^{(12)}=\lambda^{(13)}=0$,

$$\frac{\partial}{\partial\lambda^{(12)}}\gamma(b;\lambda^{(12)},\lambda^{(13)},w_1=w_2=w_3=0)|_{\lambda^{(12)}=\lambda^{(13)}=0}$$

$$= -\sqrt{\frac{\frac{3}{4}-\eta^2}{1-\eta^2}} \int_{-\infty}^{b-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon} \phi(z)\phi(\frac{\eta z}{\sqrt{1-\eta^2}})\Phi\left(\frac{\eta z\sqrt{1-\eta^2}-(\frac{1}{2}-\eta^2)\left(\frac{\eta z}{\sqrt{1-\eta^2}}\right)}{\sqrt{\frac{3}{4}-\eta^2}}\right)dz$$

$$+\sqrt{\frac{\frac{3}{4}-\eta^2}{1-\eta^2}} \int_{-\infty}^{b-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon} \phi(z)\phi(\frac{\eta z}{\sqrt{1-\eta^2}})\Phi\left(\frac{\eta z\sqrt{1-\eta^2}-(\frac{1}{2}-\eta^2)\left(\frac{\eta z}{\sqrt{1-\eta^2}}\right)}{\sqrt{\frac{3}{4}-\eta^2}}\right)dz$$

$$+\sqrt{1-\eta^2}\int_{-\infty}^{b-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon}\phi(z)\int_{-\infty}^{\frac{\eta z}{\sqrt{1-\eta^2}}}\phi(s)\phi\left(\frac{\eta z\sqrt{1-\eta^2}-(\frac{1}{2}-\eta^2)s}{\sqrt{\frac{3}{4}-\eta^2}}\right)dsdz$$

$$-\sqrt{1-\eta^2}\int_{-\infty}^{b-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon}\phi(z)\int_{-\infty}^{\frac{\eta z}{\sqrt{1-\eta^2}}}\phi(s)\phi\left(\frac{\eta z\sqrt{1-\eta^2}-(\frac{1}{2}-\eta^2)s}{\sqrt{\frac{3}{4}-\eta^2}}\right)dsdz,$$

let

$$a=\frac{\eta z}{\sqrt{\frac{\frac{3}{4}-\eta^2}{(1-\eta^2)}}}, b=-\frac{\frac{1}{2}-\eta^2}{\sqrt{\frac{3}{4}-\eta^2}}.$$

$$1+b^2=1+\frac{\left(\frac{1}{2}-\eta^2\right)^2}{\frac{3}{4}-\eta^2}=\frac{\frac{3}{4}-\eta^2+\frac{1}{4}-\eta^2+\eta^4}{\frac{3}{4}-\eta^2}=\frac{\eta^4-2\eta^2+1}{\frac{3}{4}-\eta^2}=\frac{\left(1-\eta^2\right)^2}{\frac{3}{4}-\eta^2}.$$

$$\sqrt{1+b^2} = \frac{\left(1-\eta^2\right)}{\sqrt{\frac{3}{4}-\eta^2}}.$$

$$\frac{ab}{\sqrt{1+b^2}} = \frac{\eta z}{\sqrt{\frac{\frac{3}{4}-\eta^2}{(1-\eta^2)}}} \frac{\rho\sqrt{1-\eta^2}}{\sqrt{\frac{\frac{3}{4}-\eta^2}{(1-\eta^2)}}} \frac{\sqrt{\frac{3}{4}-\eta^2}}{(1-\eta^2)} = -\frac{\left(1-\eta^2\right)}{\frac{3}{4}-\eta^2}\eta z\sqrt{\frac{\frac{3}{4}-\eta^2}{1-\eta^2}}\left(\frac{\frac{1}{2}-\eta^2}{1-\eta^2}\right)$$

$$= -\frac{\eta z\left(\frac{1}{2}-\eta^2\right)}{\sqrt{\left(\frac{3}{4}-\eta^2\right)(1-\eta^2)}}$$

$$\frac{a}{\sqrt{1+b^2}} = \frac{\eta z}{\sqrt{\frac{\frac{3}{4}-\eta^2}{(1-\eta^2)}}} \frac{\sqrt{\frac{3}{4}-\eta^2}}{(1-\eta^2)} = \frac{\eta z}{\sqrt{(1-\eta^2)}.}$$

by Lemma A.1, we have

$$\int_{-\infty}^{\frac{\eta z}{\sqrt{1-\eta^2}}} \phi(s)\phi(\frac{\eta z\sqrt{1-\eta^2}-(\frac{1}{2}-\eta^2)s}{\sqrt{\frac{3}{4}-\eta^2}})ds$$

$$= \frac{\sqrt{\frac{3}{4}-\eta^2}}{(1-\eta^2)}\phi(\frac{\eta z}{\sqrt{(1-\eta^2)}})\Phi\left(\frac{\eta z}{\sqrt{1-\eta^2}}\frac{\left(1-\eta^2\right)}{\sqrt{\frac{3}{4}-\eta^2}} - \frac{\eta z\left(\frac{1}{2}-\eta^2\right)}{\sqrt{\left(\frac{3}{4}-\eta^2\right)(1-\eta^2)}}\right).$$

Therefore,

$$\frac{\partial}{\partial\lambda^{(12)}}\gamma(b;\lambda^{(12)},\lambda^{(13)},w_1=w_2=w_3=0)|_{\lambda^{(12)}=\lambda^{(13)}=0}$$

$$= -\sqrt{\frac{\frac{3}{4}-\eta^2}{1-\eta^2}}\int_{-\infty}^{b-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon}\phi(z)\phi(\frac{\eta z}{\sqrt{1-\eta^2}})\Phi\left(\frac{\eta z\sqrt{1-\eta^2}-(\frac{1}{2}-\eta^2)\left(\frac{\eta z}{\sqrt{1-\eta^2}}\right)}{\sqrt{\frac{3}{4}-\eta^2}}\right)dz$$

$$+\sqrt{\frac{\frac{3}{4}-\eta^2}{1-\eta^2}}\int_{-\infty}^{b-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon}\phi(z)\phi(\frac{\eta z}{\sqrt{1-\eta^2}})\Phi\left(\frac{\eta z\sqrt{1-\eta^2}-(\frac{1}{2}-\eta^2)\left(\frac{\eta z}{\sqrt{1-\eta^2}}\right)}{\sqrt{\frac{3}{4}-\eta^2}}\right)dz$$

$$+\sqrt{\frac{\frac{3}{4}-\eta^2}{1-\eta^2}}\int_{-\infty}^{b-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon}\phi(z)\phi(\frac{\eta z}{\sqrt{1-\eta^2}})\Phi\left(\frac{\eta z\sqrt{1-\eta^2}-(\frac{1}{2}-\eta^2)\left(\frac{\eta z}{\sqrt{1-\eta^2}}\right)}{\sqrt{\frac{3}{4}-\eta^2}}\right)dz$$

$$-\sqrt{\frac{\frac{3}{4}-\eta^2}{1-\eta^2}}\int_{-\infty}^{b-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon}\phi(z)\phi(\frac{\eta z}{\sqrt{1-\eta^2}})\Phi\left(\frac{\eta z\sqrt{1-\eta^2}-(\frac{1}{2}-\eta^2)\left(\frac{\eta z}{\sqrt{1-\eta^2}}\right)}{\sqrt{\frac{3}{4}-\eta^2}}\right)dz$$

$$= 0.$$

Similarly, taking the partial derivative with respect to $\lambda^{(13)}$ in $\gamma$ when $w_1 = w_2 = w_3 = 0$, we have

$$
\begin{aligned}
&\frac{\partial}{\partial\lambda^{(13)}}\gamma(b;\lambda^{(12)},\lambda^{(13)},w_1=w_2=w_3=0)\\
&=-\sqrt{1-\eta^2}\int_{-\infty}^{b-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon}\phi(z)\int_{-\infty}^{\frac{\lambda^{(12)}+\eta z}{\sqrt{1-\eta^2}}}\phi(s)\phi(\frac{\left(\lambda^{(13)}+\eta z\right)\sqrt{1-\eta^2}-(\frac{1}{2}-\eta^2)s}{\sqrt{\frac{3}{4}-\eta^2}})dsdz\\
&\quad-\sqrt{1-\eta^2}\int_{-\infty}^{b-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon}\phi(z)\int_{-\infty}^{\frac{-\lambda^{(12)}+\eta z}{\sqrt{1-\eta^2}}}\phi(s)\phi(\frac{\left(\lambda^{(13)}-\lambda^{(12)}+\eta z\right)\sqrt{1-\eta^2}-(\frac{1}{2}-\eta^2)s}{\sqrt{\frac{3}{4}-\eta^2}})dsdz\\
&\quad+\sqrt{\frac{\frac{3}{4}-\eta^2}{1-\eta^2}}\int_{-\infty}^{b-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon}\phi(z)\phi(\frac{-\lambda^{(13)}+\eta z}{\sqrt{1-\eta^2}})\\
&\qquad\times\Phi\left(\frac{\left(\lambda^{(12)}-\lambda^{(13)}+\eta z\right)\sqrt{1-\eta^2}-(\frac{1}{2}-\eta^2)\left(\frac{-\lambda^{(13)}+\eta z}{\sqrt{1-\eta^2}}\right)}{\sqrt{\frac{3}{4}-\eta^2}}\right)dz\\
&\quad+\sqrt{1-\eta^2}\int_{-\infty}^{b-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon}\phi(z)\int_{-\infty}^{\frac{-\lambda^{(13)}+\eta z}{\sqrt{1-\eta^2}}}\phi(s)\phi(\frac{\left(\lambda^{(12)}-\lambda^{(13)}+\eta z\right)\sqrt{1-\eta^2}-(\frac{1}{2}-\eta^2)s}{\sqrt{\frac{3}{4}-\eta^2}})dsdz.
\end{aligned}
$$

When $\lambda^{(12)} = \lambda^{(13)} = 0$,

$$
\begin{aligned}
&\frac{\partial}{\partial\lambda^{(13)}}\gamma(b;\lambda^{(12)},\lambda^{(13)},w_1=w_2=w_3=0)|_{\lambda^{(12)}=\lambda^{(13)}=0}\\
&=-\sqrt{1-\eta^2}\int_{-\infty}^{b-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon}\phi(z)\int_{-\infty}^{\frac{\eta z}{\sqrt{1-\eta^2}}}\phi(s)\phi(\frac{\eta z\sqrt{1-\eta^2}-(\frac{1}{2}-\eta^2)s}{\sqrt{\frac{3}{4}-\eta^2}})dsdz\\
&\quad-\sqrt{1-\eta^2}\int_{-\infty}^{b-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon}\phi(z)\int_{-\infty}^{\frac{\eta z}{\sqrt{1-\eta^2}}}\phi(s)\phi(\frac{\eta z\sqrt{1-\eta^2}-(\frac{1}{2}-\eta^2)s}{\sqrt{\frac{3}{4}-\eta^2}})dsdz\\
&\quad+\sqrt{\frac{\frac{3}{4}-\eta^2}{1-\eta^2}}\int_{-\infty}^{b-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon}\phi(z)\phi(\frac{\eta z}{\sqrt{1-\eta^2}})\Phi\left(\frac{\eta z\sqrt{1-\eta^2}-(\frac{1}{2}-\eta^2)\left(\frac{\eta z}{\sqrt{1-\eta^2}}\right)}{\sqrt{\frac{3}{4}-\eta^2}}\right)dz\\
&\quad+\sqrt{1-\eta^2}\int_{-\infty}^{b-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon}\phi(z)\int_{-\infty}^{\frac{\eta z}{\sqrt{1-\eta^2}}}\phi(s)\phi(\frac{\eta z\sqrt{1-\eta^2}-(\frac{1}{2}-\eta^2)s}{\sqrt{\frac{3}{4}-\eta^2}})dsdz,
\end{aligned}
$$

let

$$a=\frac{\eta z}{\sqrt{\frac{\frac{3}{4}-\eta^2}{(1-\eta^2)}}}, b=-\frac{\frac{1}{2}-\eta^2}{\sqrt{\frac{3}{4}-\eta^2}}.$$

$$1+b^2=1+\frac{\left(\frac{1}{2}-\eta^2\right)^2}{\frac{3}{4}-\eta^2}=\frac{\frac{3}{4}-\eta^2+\frac{1}{4}-\eta^2+\eta^4}{\frac{3}{4}-\eta^2}=\frac{\eta^4-2\eta^2+1}{\frac{3}{4}-\eta^2}=\frac{\left(1-\eta^2\right)^2}{\frac{3}{4}-\eta^2}.$$

$$\sqrt{1+b^2}=\frac{\left(1-\eta^2\right)}{\sqrt{\frac{3}{4}-\eta^2}}.$$

$$\frac{ab}{\sqrt{1+b^2}}=\frac{\eta z}{\sqrt{\frac{\frac{3}{4}-\eta^2}{(1-\eta^2)}}}\frac{\rho\sqrt{1-\eta^2}}{\sqrt{\frac{\frac{3}{4}-\eta^2}{(1-\eta^2)}}}\frac{\sqrt{\frac{3}{4}-\eta^2}}{(1-\eta^2)}=-\frac{\left(1-\eta^2\right)}{\frac{3}{4}-\eta^2}\eta z\sqrt{\frac{\frac{3}{4}-\eta^2}{1-\eta^2}}\left(\frac{\frac{1}{2}-\eta^2}{1-\eta^2}\right)$$

$$=-\frac{\eta z\left(\frac{1}{2}-\eta^2\right)}{\sqrt{\left(\frac{3}{4}-\eta^2\right)\left(1-\eta^2\right)}}$$

$$\frac{a}{\sqrt{1+b^2}}=\frac{\eta z}{\sqrt{\frac{\frac{3}{4}-\eta^2}{(1-\eta^2)}}}\frac{\sqrt{\frac{3}{4}-\eta^2}}{(1-\eta^2)}=\frac{\eta z}{\sqrt{(1-\eta^2)}.}$$

by Lemma A.1, we have

$$\int_{-\infty}^{\frac{\eta z}{\sqrt{1-\eta^2}}}\phi(s)\phi(\frac{\eta z\sqrt{1-\eta^2}-(\frac{1}{2}-\eta^2)s}{\sqrt{\frac{3}{4}-\eta^2}})ds$$

$$=\frac{\sqrt{\frac{3}{4}-\eta^2}}{(1-\eta^2)}\phi(\frac{\eta z}{\sqrt{(1-\eta^2)}})\Phi\left(\frac{\eta z}{\sqrt{1-\eta^2}}\frac{\left(1-\eta^2\right)}{\sqrt{\frac{3}{4}-\eta^2}}-\frac{\eta z\left(\frac{1}{2}-\eta^2\right)}{\sqrt{\left(\frac{3}{4}-\eta^2\right)\left(1-\eta^2\right)}}\right).$$

Therefore,

$$\frac{\partial}{\partial\lambda^{(13)}}\gamma(b;\lambda^{(12)},\lambda^{(13)},w_1=w_2=w_3=0)|_{\lambda^{(12)}=\lambda^{(13)}=0}$$

$$=-\sqrt{1-\eta^2}\int_{-\infty}^{b-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon}\phi(z)\int_{-\infty}^{\frac{\eta z}{\sqrt{1-\eta^2}}}\phi(s)\phi(\frac{\eta z\sqrt{1-\eta^2}-(\frac{1}{2}-\eta^2)s}{\sqrt{\frac{3}{4}-\eta^2}})dsdz$$

$$-\sqrt{1-\eta^2}\int_{-\infty}^{b-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon}\phi(z)\int_{-\infty}^{\frac{\eta z}{\sqrt{1-\eta^2}}}\phi(s)\phi(\frac{\eta z\sqrt{1-\eta^2}-(\frac{1}{2}-\eta^2)s}{\sqrt{\frac{3}{4}-\eta^2}})dsdz$$

$$+\sqrt{\frac{\frac{3}{4}-\eta^2}{1-\eta^2}}\int_{-\infty}^{b-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon}\phi(z)\phi(\frac{\eta z}{\sqrt{1-\eta^2}})\Phi\left(\frac{\eta z\sqrt{1-\eta^2}-(\frac{1}{2}-\eta^2)\left(\frac{\eta z}{\sqrt{1-\eta^2}}\right)}{\sqrt{\frac{3}{4}-\eta^2}}\right)dz$$

$$
\begin{aligned}
&+\sqrt{1-\eta^2}\int_{-\infty}^{b-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon}\phi(z)\int_{-\infty}^{\frac{\eta z}{\sqrt{1-\eta^2}}}\phi(s)\phi(\frac{\eta z\sqrt{1-\eta^2}-(\frac{1}{2}-\eta^2)s}{\sqrt{\frac{3}{4}-\eta^2}})dsdz\\
=&-\sqrt{1-\eta^2}\int_{-\infty}^{b-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon}\phi(z)\int_{-\infty}^{\frac{\eta z}{\sqrt{1-\eta^2}}}\phi(s)\phi(\frac{\eta z\sqrt{1-\eta^2}-(\frac{1}{2}-\eta^2)s}{\sqrt{\frac{3}{4}-\eta^2}})dsdz\\
&+\sqrt{1-\eta^2}\int_{-\infty}^{b-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon}\phi(z)\int_{-\infty}^{\frac{\eta z}{\sqrt{1-\eta^2}}}\phi(s)\phi(\frac{\eta z\sqrt{1-\eta^2}-(\frac{1}{2}-\eta^2)s}{\sqrt{\frac{3}{4}-\eta^2}})dsdz\\
&-\sqrt{\frac{\frac{3}{4}-\eta^2}{1-\eta^2}}\int_{-\infty}^{b-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon}\phi(z)\phi(\frac{\eta z}{\sqrt{1-\eta^2}})\Phi\left(\frac{\eta z\sqrt{1-\eta^2}}{\sqrt{\frac{3}{4}-\eta^2}}-\frac{\eta z\left(\frac{1}{2}-\eta^2\right)}{\sqrt{\left(\frac{3}{4}-\eta^2\right)\left(1-\eta^2\right)}}\right)dz\\
&+\sqrt{\frac{\frac{3}{4}-\eta^2}{1-\eta^2}}\int_{-\infty}^{b-\sqrt{\frac{n}{2\sigma_Y^2}}\epsilon}\phi(z)\phi(\frac{\eta z}{\sqrt{1-\eta^2}})\Phi\left(\frac{\eta z\sqrt{1-\eta^2}-(\frac{1}{2}-\eta^2)\left(\frac{\eta z}{\sqrt{1-\eta^2}}\right)}{\sqrt{\frac{3}{4}-\eta^2}}\right)dz\\
=&\,0.
\end{aligned}
$$

This implies that given $b$, $\lambda^{(12)}=\lambda^{(13)}=0$ is the critical point of $\gamma(b;\lambda^{(12)},\lambda^{(13)},w_1,w_2,w_3)$ when $w_1=w_2=w_3=0$. The proof of uniqueness of the critical point is not easy, through numeric justification, Figure C.1 and C.2 demonstrate that $\lambda^{(12)}=\lambda^{(13)}=0$ is the unique critical point under $H_0$.

□

## C.4 Numeric Justification for Property 4.3.3

We use numerical justification to show that given $b$, $\gamma(b;w_1,w_2,w_3)$ is monotonically increasing with $w_1,w_2,w_3$ at $\alpha=0.025,0.05$. Figure C.3 and Figure C.4 show the $Pr(W>b)=0.025$ and the $Pr(W>b)=0.05$ versus $w_2$ and $w_3$ by $w_1$ and $\tau$ when the interim and final endpoints are the same. As seen in the figures, under $H_0$, the $Pr(W>b)$ is monotonically increasing with $w_1,w_2$ and $w_3$.

Figure C.1: $Pr(W > b) = 0.025$ versus $\lambda^{(12)}$ and $\lambda^{(13)}$ when $w_1 = w_2 = w_3 = 0$ by $\tau$ and $\eta$ when different endpoints at interim and final analyses.

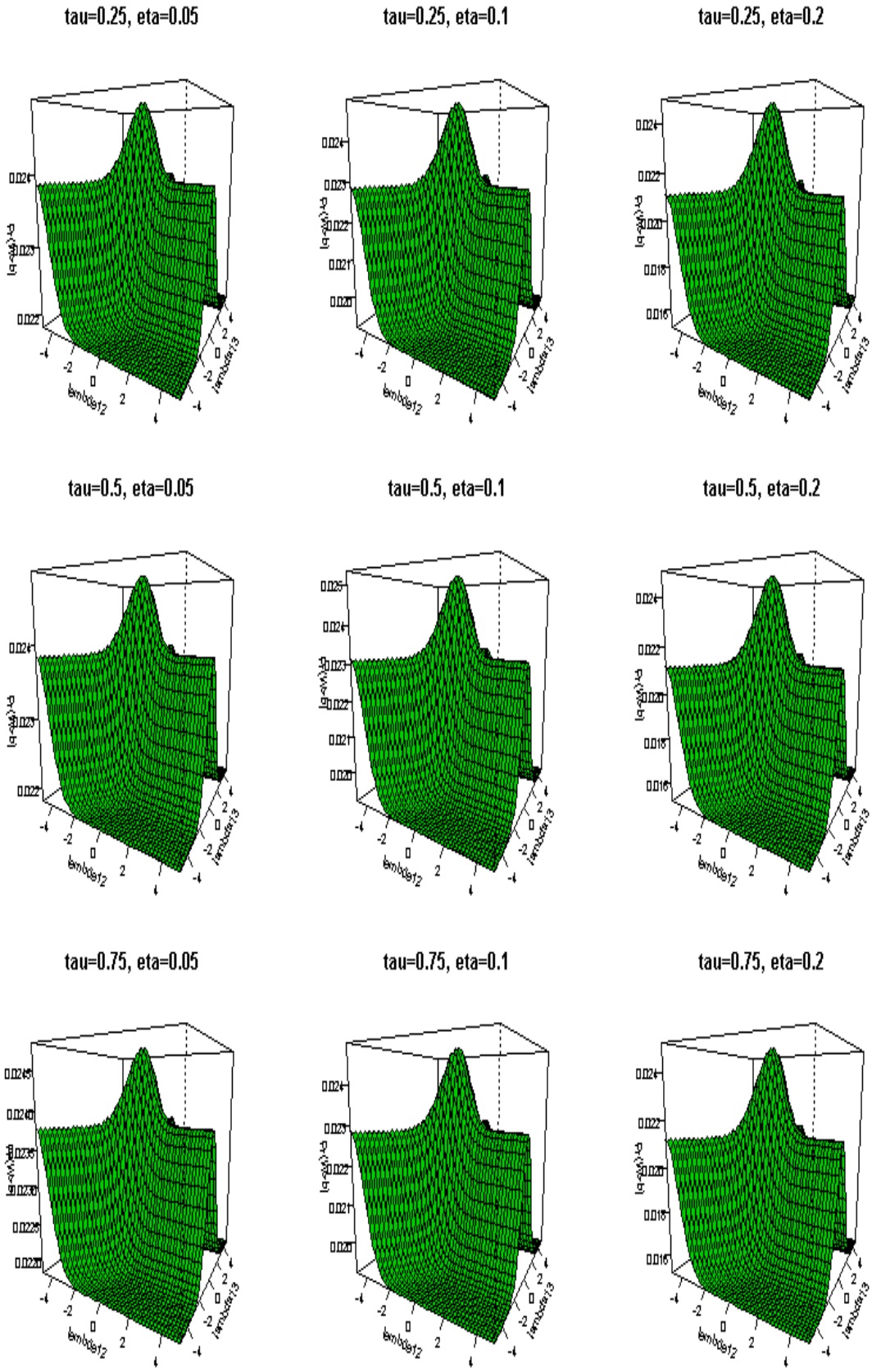


Figure C.2: $Pr(W > b) = 0.05$ versus $\lambda^{(12)}$ and $\lambda^{(13)}$ when $w_1 = w_2 = w_3 = 0$ by $\tau$ and $\eta$ when different endpoints at interim and final analyses.

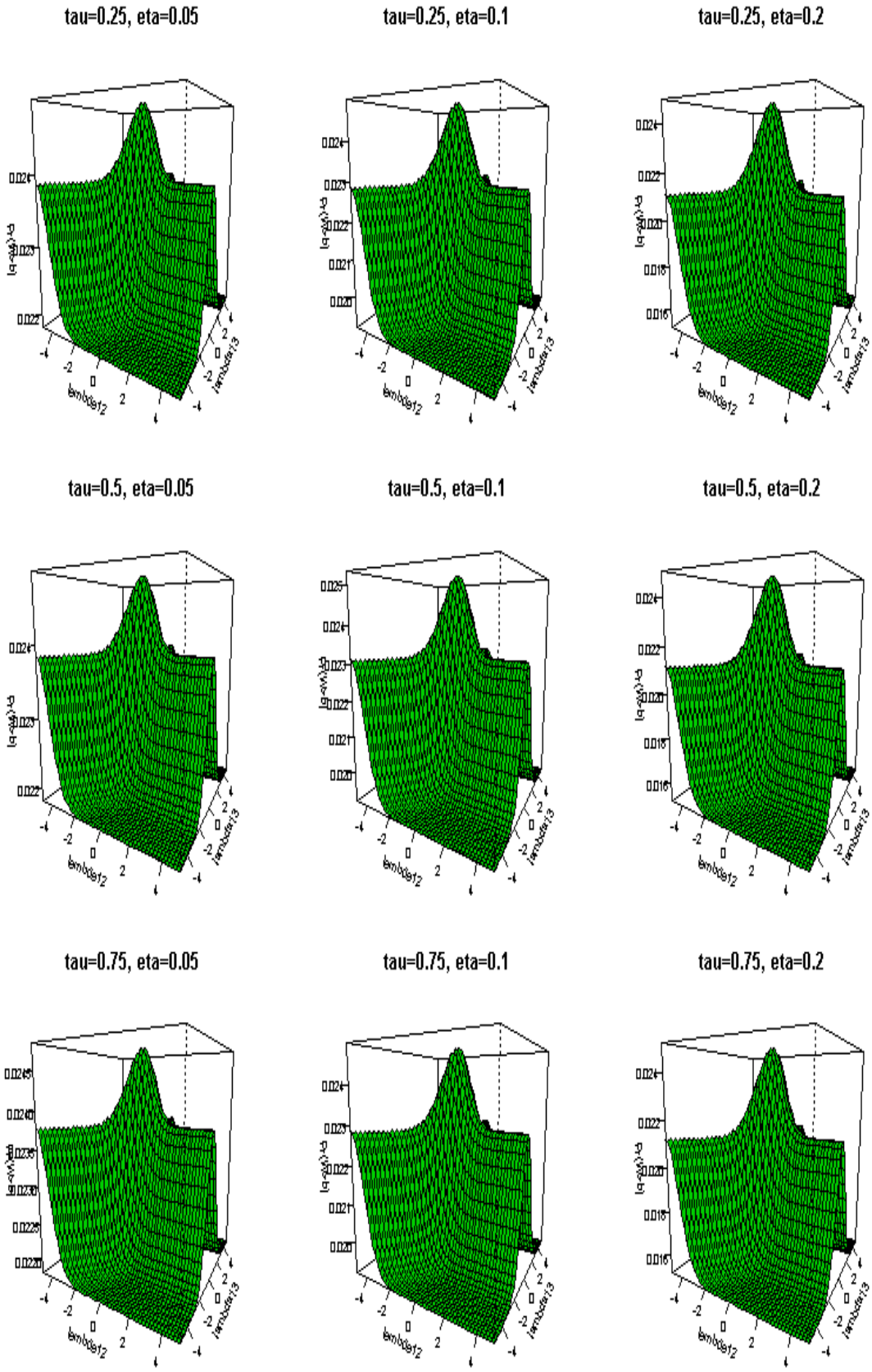


Figure C.3: $Pr(W > b) = 0.025$ versus $w_2$, and $w_3$ by $w_1$ and $\tau$ when same endpoints at interim and final analyses.

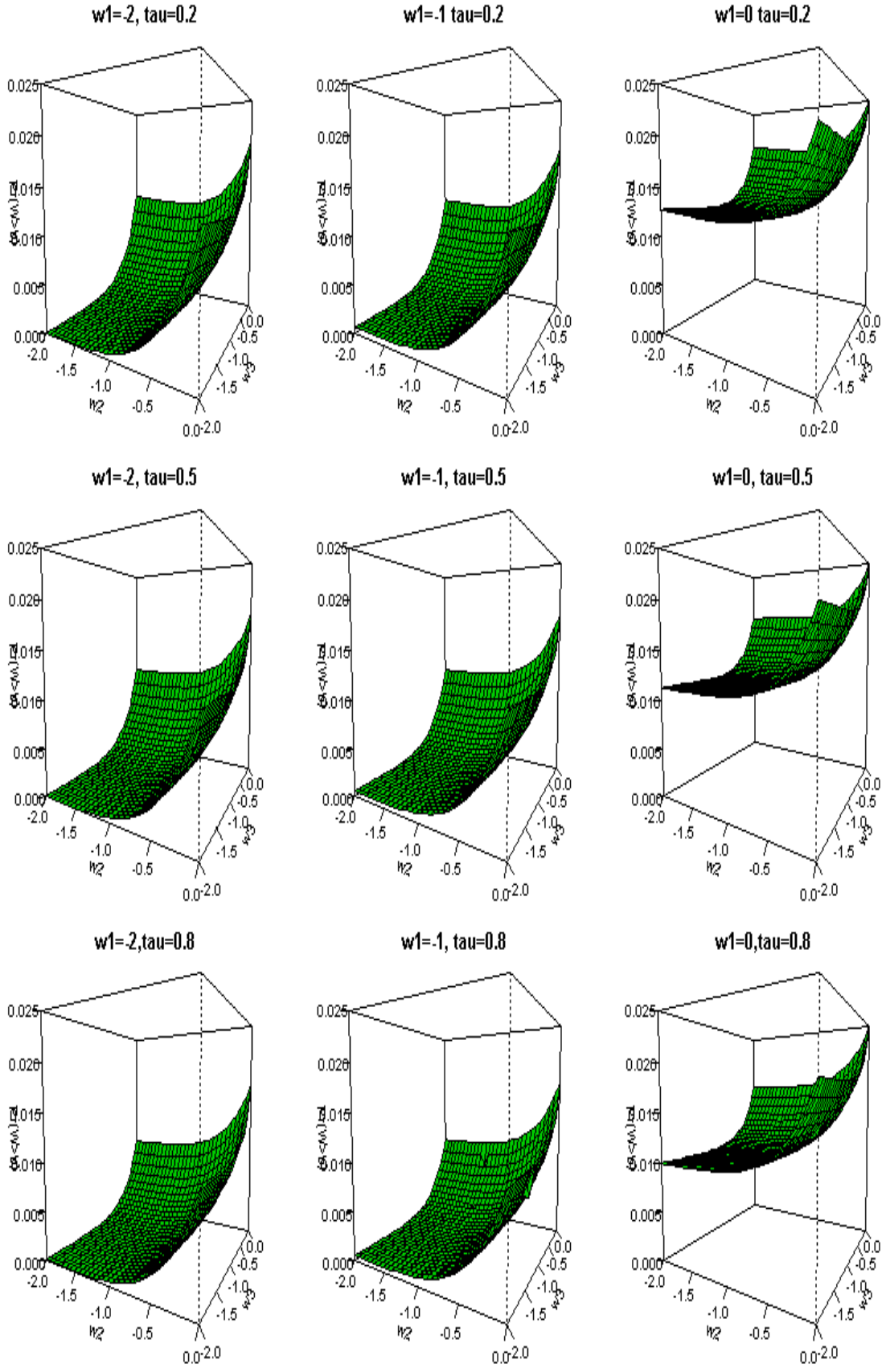
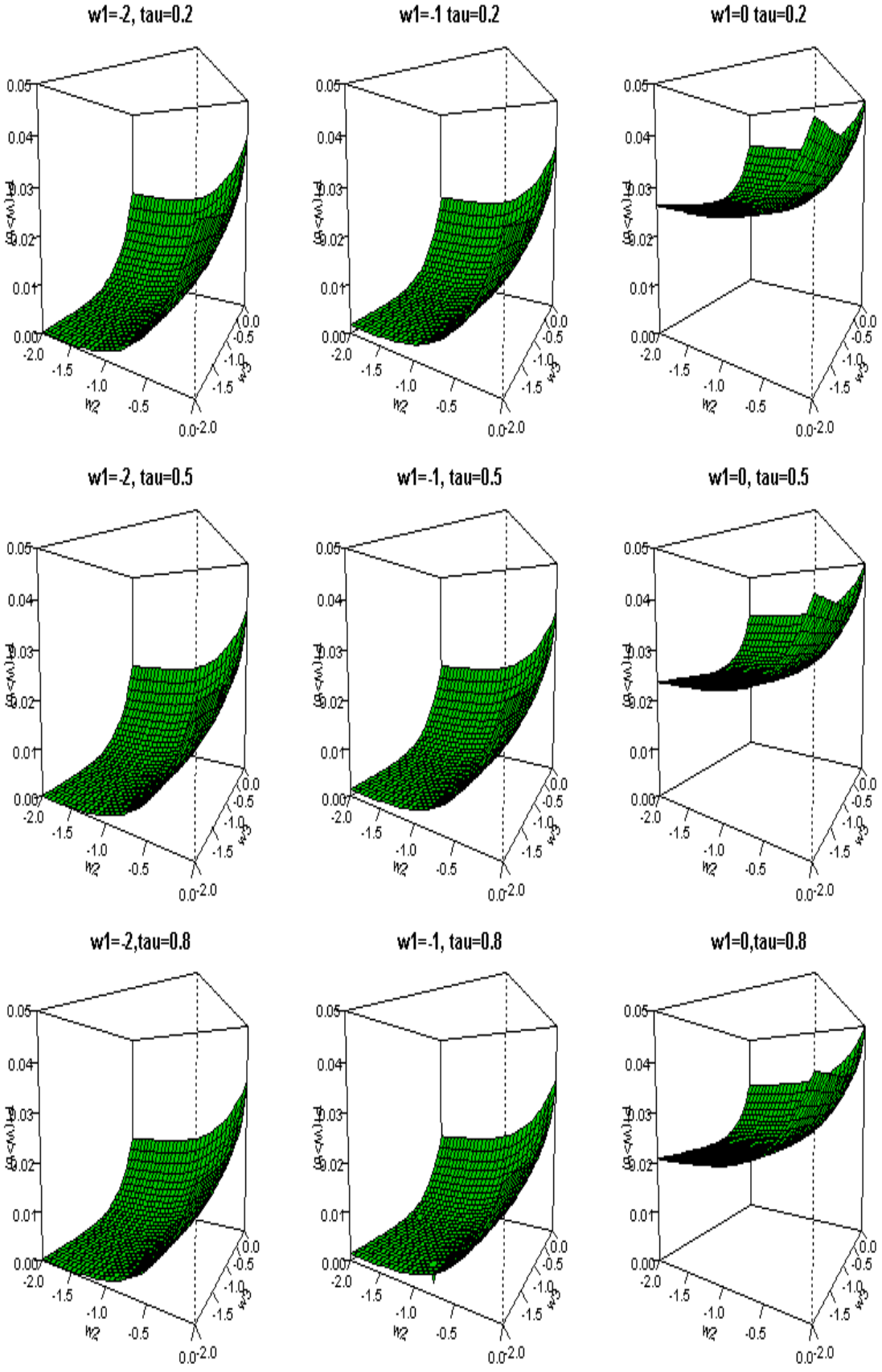


Figure C.4: $Pr(W > b) = 0.05$ versus $w_2$, and $w_3$ by $w_1$ and $\tau$ when same endpoints at interim and final analyses .



# Appendix D

# Mathematica Program for Solving the Non-linear Equations for Normal Approximation of the Sample Size

## D.1 Pre-determine the Information Time $\tau$ when Different Endpoints at Interim and Final Analyses

```
NIapproximation[NItau_,
                NIrho_,
                NIalpha_,
                NIsigmaX_,
                NIsigmaY_,
                NInu12_,
                NIbeta_,
                NIdelta1_,
                NIdelta2_,
                NIepsilon_,
                NInInitial_,
                NIbeta1Initial_,
                NIbeta2Initial_
                ] :=
  Module[{NIeta,
          NIsqrt2pi,
          NI2sigmaX2,
          NI2sigmaY2,
          NInu12square,
          NIzalpha,
          NIPvalue,
          NIequation0Value,
```

```
        NIequation1Value,
        NIequation2Value
        },
(
  NIeta = Sqrt[NItau]/2*NIrho;
  NIsqrt2pi = N[Sqrt[2*Pi]];
  NI2sigmaX2 = 2*NIsigmaX*NIsigmaX;
  NI2sigmaY2 = 2*NIsigmaY*NIsigmaY;
  NInu12square = NInu12^2;
  NIzalpha = Quantile[NormalDistribution[0, 1], 1-NIalpha];
  NIlambda[n_] :=
    Sqrt[n * NItau / NI2sigmaX2] * NInu12;
  NILambda[n_] :=
    NIeta / NIsqrt2pi * Exp[-NIlambda[n]^2/2];
  NIp[n_] :=
    CDF[NormalDistribution[0, 1], NIlambda[n]];
  NImu0 = Sqrt[2/Pi]*NIeta;
  NIsigma0 = Sqrt[1 - 2/Pi*NIeta*NIeta];
  NImu1[n_] :=
    NILambda[n] / NIp[n];
  NIsigma1[n_] :=
    Sqrt[1 - NIlambda[n]*NIeta*NImu1[n] - NImu1[n]^2];
  NImu2[n_] :=
    NILambda[n] / (1-NIp[n]);
  NIsigma2[n_] :=
    Sqrt[1 + NIlambda[n]*NIeta*NImu2[n] - NImu2[n]^2];
  NIequation0[n_, beta1_, beta2_] :=
    Module[{rv},
      (rv = N[NIp[n] * beta1 + (1 - NIp[n]) * beta2 - (1-NIbeta)];
       Return[rv]
       )
    ];
  NIequation1[n_, beta1_] :=
    Module[{rv},
```

```
   (rv =
      Quantile[NormalDistribution[0, 1], 1-beta1] * NIsigma1[n] +
      NIzalpha * NIsigma0 +
      NImu0 - NImu1[n] - Sqrt[n/2/NIsigmaY^2] * (NIdelta1 + NIepsilon);
    Return[rv]
    )
   ];
NIequation2[n_, beta2_] :=
  Module[{rv},
   (rv =
      Quantile[NormalDistribution[0, 1], 1-beta2] * NIsigma2[n] +
      NIzalpha * NIsigma0 +
      NImu0 - NImu2[n] - Sqrt[n/2/NIsigmaY^2] * (NIdelta2 + NIepsilon);
    Return[rv]
    )
   ];
NIsolution =
  N[FindRoot[{NIequation0[nx, beta1x, beta2x],
              NIequation1[nx, beta1x],
              NIequation2[nx, beta2x]
              },
             {{nx, NInInitial},
              {beta1x, NIbeta1Initial},
              {beta2x, NIbeta2Initial}
              },
             MaxIterations -> 100000
             ]];
NIequation0Value =
 N[NIequation0[nx/.NIsolution, beta1x/.NIsolution, beta2x/.NIsolution]];
NIequation1Value =
 N[NIequation1[nx/.NIsolution, beta1x/.NIsolution]];
NIequation2Value =
 N[NIequation2[nx/.NIsolution, beta2x/.NIsolution]];
NIPvalue = N[NIp[nx /. NIsolution]];
```

```
    Return[Join[{tau -> NItau,
                 rho -> NIrho,
                 alpha -> NIalpha,
                 sigmaX -> NIsigmaX,
                 sigmaY -> NIsigmaY,
                 nu12 -> NInu12,
                 beta -> NIbeta,
                 delta1 -> NIdelta1,
                 delta2 -> NIdelta2,
                 epsilon -> NIepsilon,
                 e0 -> NIequation0Value,
                 e1 -> NIequation1Value,
                 e2 -> NIequation2Value,
                 p -> NIPvalue,
                 },
                NIsolution
                ]
           ]
  )
  ]
```

## D.2 Pre-determine the Winning Probability $p$ when Different Endpoints at Interim and Final Analyses

```
NIapproximation[NIp_,
                NIrho_,
                NIalpha_,
                NIsigmaX_,
                NIsigmaY_,
                NInu12_,
                NIbeta_,
                NIdelta1_,
                NIdelta2_,
                NIepsilon_,
```

```
                NInInitial_,
                NIbeta1Initial_,
                NIbeta2Initial_
                ] :=
Module[{NIzp,
        NIn1,
        NIlambda,
        NIsqrt2pi,
        NI2sigmaX2,
        NI2sigmaY2,
        NInu12square,
        NIzalpha,
        NIequation0Value,
        NIequation1Value,
        NIequation2Value
        },
(
  NIsqrt2pi = N[Sqrt[2*Pi]];
  NI2sigmaX2 = 2*NIsigmaX*NIsigmaX;
  NI2sigmaY2 = 2*NIsigmaY*NIsigmaY;
  NInu12square = NInu12^2;
  NIzalpha = Quantile[NormalDistribution[0, 1], 1-NIalpha];
  NIzp =
    Quantile[NormalDistribution[0, 1], 1-NIp];
  NIn1 =
    NI2sigmaY2 * NIzp^2 / NInu12square;
  NItau[n_] :=
    NIn1 / n;
  NIeta[n_] := Sqrt[NItau[n]]/2*NIrho;
  NIlambda =
    Sqrt[NIn1 / NI2sigmaY2] * NInu12;
  NILambda[n_] :=
    NIeta[n] / NIsqrt2pi * Exp[-NIlambda^2/2];
  NImu0[n_] :=
```

```
  Sqrt[2/Pi]*NIeta[n];
NIsigma0[n_] :=
  Sqrt[1 - 2/Pi*NIeta[n]*NIeta[n]];
NImu1[n_] :=
  NILambda[n] / NIp;
NIsigma1[n_] :=
  Sqrt[1 - NIlambda*NIeta[n]*NImu1[n] - NImu1[n]^2];
NImu2[n_] :=
  NILambda[n] / (1-NIp);
NIsigma2[n_] :=
  Sqrt[1 + NIlambda*NIeta[n]*NImu2[n] - NImu2[n]^2];
NIequation0[n_, beta1_, beta2_] :=
  Module[{rv},
    (rv = N[NIp * beta1 + (1 - NIp) * beta2 - (1-NIbeta)];
     Return[rv]
     )
  ];
NIequation1[n_, beta1_] :=
  Module[{rv},
   (rv =
      Quantile[NormalDistribution[0, 1], 1-beta1] * NIsigma1[n] +
      NIzalpha * NIsigma0[n] +
      NImu0[n] - NImu1[n] - Sqrt[n/2/NIsigmaY^2] * (NIdelta1 + NIepsilon);
    Return[rv]
    )
   ];
NIequation2[n_, beta2_] :=
  Module[{rv},
   (rv =
      Quantile[NormalDistribution[0, 1], 1-beta2] * NIsigma2[n] +
      NIzalpha * NIsigma0[n] +
      NImu0[n] - NImu2[n] - Sqrt[n/2/NIsigmaY^2] * (NIdelta2 + NIepsilon);
    Return[rv]
    )
```

```
   ];
NIsolution =
  N[FindRoot[{NIequation0[nx, beta1x, beta2x],
              NIequation1[nx, beta1x],
              NIequation2[nx, beta2x]
              },
             {{nx, NInInitial},
              {beta1x, NIbeta1Initial},
              {beta2x, NIbeta2Initial}
              },
             MaxIterations -> 1000000
             ]];
NIequation0Value =
 N[NIequation0[nx/.NIsolution, beta1x/.NIsolution, beta2x/.NIsolution]];
NIequation1Value =
 N[NIequation1[nx/.NIsolution, beta1x/.NIsolution]];
NIequation2Value =
 N[NIequation2[nx/.NIsolution, beta2x/.NIsolution]];
NItauValue = N[NItau[nx /. NIsolution]];
Return[Join[{p -> NIp,
             rho -> NIrho,
             alpha -> NIalpha,
             sigmaX -> NIsigmaX,
             sigmaY -> NIsigmaY,
             nu12 -> NInu12,
             beta -> NIbeta,
             delta1 -> NIdelta1,
             delta2 -> NIdelta2,
             epsilon -> NIepsilon,
             e0 -> NIequation0Value,
             e1 -> NIequation1Value,
             e2 -> NIequation2Value,
             tau -> NItauValue,
             zp -> NIzp,
```

```
                n1 -> NIn1,
                lambda -> NIlambda
                },
               NIsolution
               ]
          ]
  )
  ]
```

## D.3 Pre-determine the Information Time $\tau$ when Same Endpoints at Interim and Final Analyses

```
NIapproximation[NItau_,
               NIrho_,
               NIalpha_,
               NIsigmaY_,
               NIbeta_,
               NIdelta1_,
               NIdelta2_,
               NIepsilon_,
               NInInitial_,
               NIbeta1Initial_,
               NIbeta2Initial_
               ] :=
  Module[{NIeta,
          NIsqrt2pi,
          NI2sigmaY2,
          NIzalpha,
          NIPvalue,
          NIequation0Value,
          NIequation1Value,
          NIequation2Value
          },
  (
```

```
NIeta = Sqrt[NItau]/2*NIrho;
NIsqrt2pi = N[Sqrt[2*Pi]];
NI2sigmaY2 = 2*NIsigmaY*NIsigmaY;
NIzalpha = Quantile[NormalDistribution[0, 1], 1-NIalpha];
NIlambda[n_] :=
  Sqrt[n * NItau / NI2sigmaY2] * (NIdelta1-NIdelta2);
NILambda[n_] :=
  NIeta / NIsqrt2pi * Exp[-NIlambda[n]^2/2];
NIp[n_] :=
  CDF[NormalDistribution[0, 1], NIlambda[n]];
NImu0 = Sqrt[2/Pi]*NIeta;
NIsigma0 = Sqrt[1 - 2/Pi*NIeta*NIeta];
NImu1[n_] :=
  NILambda[n] / NIp[n];
NIsigma1[n_] :=
  Sqrt[1 - NIlambda[n]*NIeta*NImu1[n] - NImu1[n]^2];
NImu2[n_] :=
  NILambda[n] / (1-NIp[n]);
NIsigma2[n_] :=
  Sqrt[1 + NIlambda[n]*NIeta*NImu2[n] - NImu2[n]^2];
NIequation0[n_, beta1_, beta2_] :=
  Module[{rv},
    (rv = N[NIp[n] * beta1 + (1 - NIp[n]) * beta2 - (1-NIbeta)];
     Return[rv]
     )
  ];
NIequation1[n_, beta1_] :=
  Module[{rv},
   (rv =
      Quantile[NormalDistribution[0, 1], 1-beta1] * NIsigma1[n] +
      NIzalpha * NIsigma0 +
      NImu0 - NImu1[n] - Sqrt[n/2/NIsigmaY^2] * (NIdelta1 + NIepsilon);
    (*Print["NIequation1.rv=", rv];*)
    Return[rv]
```

```
    )
   ];
NIequation2[n_, beta2_] :=
  Module[{rv},
   (rv =
      Quantile[NormalDistribution[0, 1], 1-beta2] * NIsigma2[n] +
      NIzalpha * NIsigma0 +
      NImu0 - NImu2[n] - Sqrt[n/2/NIsigmaY^2] * (NIdelta2 + NIepsilon);
    (*Print["NIequation2.rv=", rv];*)
    Return[rv]
    )
   ];
NIsolution =
  N[FindRoot[{NIequation0[nx, beta1x, beta2x],
              NIequation1[nx, beta1x],
              NIequation2[nx, beta2x]
              },
             {{nx, NInInitial},
              {beta1x, NIbeta1Initial},
              {beta2x, NIbeta2Initial}
              },
             MaxIterations -> 100000
             ]];
NIequation0Value =
 N[NIequation0[nx/.NIsolution, beta1x/.NIsolution, beta2x/.NIsolution]];
NIequation1Value =
 N[NIequation1[nx/.NIsolution, beta1x/.NIsolution]];
NIequation2Value =
 N[NIequation2[nx/.NIsolution, beta2x/.NIsolution]];
NIPvalue = N[NIp[nx /. NIsolution]];
Return[Join[{tau -> NItau,
             rho -> NIrho,
             alpha -> NIalpha,
             sigmaY -> NIsigmaY,
```

```
                beta -> NIbeta,
                delta1 -> NIdelta1,
                delta2 -> NIdelta2,
                epsilon -> NIepsilon,
                e0 -> NIequation0Value,
                e1 -> NIequation1Value,
                e2 -> NIequation2Value,
                p -> NIPvalue,
                },
              NIsolution
              ]
         ]
  )
  ]
```

## D.4 Pre-determine the Winning Probability $p$ when Same Endpoints at Interim and Final Analyses

```
NIapproximation[NIp_,
                NIrho_,
                NIalpha_,
                NIsigmaY_,
                NIbeta_,
                NIdelta1_,
                NIdelta2_,
                NIepsilon_,
                NInInitial_,
                NIbeta1Initial_,
                NIbeta2Initial_
                ] :=
  Module[{NIzp,
          NIn1,
          NIlambda,
          NIsqrt2pi,
```

```
        NI2sigmaY2,
        NIzalpha,
        NIequation0Value,
        NIequation1Value,
        NIequation2Value
        },
(
  NIsqrt2pi = N[Sqrt[2*Pi]];
  NI2sigmaY2 = 2*NIsigmaY*NIsigmaY;
  NIzalpha = Quantile[NormalDistribution[0, 1], 1-NIalpha];
  NIzp =
    Quantile[NormalDistribution[0, 1], 1-NIp];
  NIn1 =
    NI2sigmaY2 * NIzp^2 / (NIdelta1-NIdelta2)^2;
  NItau[n_] :=
    NIn1 / n;
  NIeta[n_] := Sqrt[NItau[n]]/2*NIrho;
  NIlambda =
    Sqrt[NIn1 / NI2sigmaY2] * (NIdelta1-NIdelta2);
  NILambda[n_] :=
    NIeta[n] / NIsqrt2pi * Exp[-NIlambda^2/2];
  NImu0[n_] :=
    Sqrt[2/Pi]*NIeta[n];
  NIsigma0[n_] :=
    Sqrt[1 - 2/Pi*NIeta[n]*NIeta[n]];
  NImu1[n_] :=
    NILambda[n] / NIp;
  NIsigma1[n_] :=
    Sqrt[1 - NIlambda*NIeta[n]*NImu1[n] - NImu1[n]^2];
  NImu2[n_] :=
    NILambda[n] / (1-NIp);
  NIsigma2[n_] :=
    Sqrt[1 + NIlambda*NIeta[n]*NImu2[n] - NImu2[n]^2];
  NIequation0[n_, beta1_, beta2_] :=
```

```
  Module[{rv},
    (rv = N[NIp * beta1 + (1 - NIp) * beta2 - (1-NIbeta)];
     Return[rv]
     )
  ];
NIequation1[n_, beta1_] :=
  Module[{rv},
   (rv =
      Quantile[NormalDistribution[0, 1], 1-beta1] * NIsigma1[n] +
      NIzalpha * NIsigma0[n] +
      NImu0[n] - NImu1[n] - Sqrt[n/2/NIsigmaY^2] * (NIdelta1 + NIepsilon);
    Return[rv]
    )
   ];
NIequation2[n_, beta2_] :=
  Module[{rv},
   (rv =
      Quantile[NormalDistribution[0, 1], 1-beta2] * NIsigma2[n] +
      NIzalpha * NIsigma0[n] +
      NImu0[n] - NImu2[n] - Sqrt[n/2/NIsigmaY^2] * (NIdelta2 + NIepsilon);
    Return[rv]
    )
   ];
NIsolution =
  N[FindRoot[{NIequation0[nx, beta1x, beta2x],
              NIequation1[nx, beta1x],
              NIequation2[nx, beta2x]
              },
             {{nx, NInInitial},
              {beta1x, NIbeta1Initial},
              {beta2x, NIbeta2Initial}
              },
             MaxIterations -> 1000000
             ]];
```

```
  NIequation0Value =
   N[NIequation0[nx/.NIsolution, beta1x/.NIsolution, beta2x/.NIsolution]];
  NIequation1Value =
   N[NIequation1[nx/.NIsolution, beta1x/.NIsolution]];
  NIequation2Value =
   N[NIequation2[nx/.NIsolution, beta2x/.NIsolution]];
  NItauValue = N[NItau[nx /. NIsolution]];
  Return[Join[{p -> NIp,
               rho -> NIrho,
               alpha -> NIalpha,
               sigmaY -> NIsigmaY,
               beta -> NIbeta,
               delta1 -> NIdelta1,
               delta2 -> NIdelta2,
               epsilon -> NIepsilon,
               e0 -> NIequation0Value,
               e1 -> NIequation1Value,
               e2 -> NIequation2Value,
               tau -> NItauValue,
               zp -> NIzp,
               n1 -> NIn1,
               lambda -> NIlambda
               },
              NIsolution
              ]
         ]
)
]
```

# Appendix E

## R Program for determine the critical value, sample size and power for the test statistic

```
#=====================================================================
#Find the Critical-value for non-inferiority test (surrogate endpoints
#used)-for two-stage winner design with two experimental treatments and
#an active control. Function findCutOffw12_eta_cstar is the function
#to determine the critical value for a two-stage winner design with
#one interim treatment selection, input parameters are:
#cstarBegin, cstarEnd(range to search the critical value), alpha
#(type I error), and eta(sqrt(tau)/2*rho).
#=====================================================================
gamma_w1w2_eta_cstar <- function(cstar,eta) {
  #---------------------------------------------------------------------
  sOneEta2 <- sqrt(1-eta^2)
  k<-eta/sOneEta2
  #---------------------------------------------------------------------
  term1 <- integrate(function(z1e) {
      d1 <- dnorm(z1e)
      d3 <- pnorm(k*z1e)
      rvTerm1Function <- d1*d3
      return(rvTerm1Function)
    }
    , -Inf, cstar)$value
  #---------------------------------------------------------------------
  return(1-2*term1)
  #---------------------------------------------------------------------
}
```

```
findCutOffw12_eta_cstar<- function(cstarBegin, cstarEnd, alpha,eta)
{
  return(uniroot(function(cstar) { return(alpha -
  gamma_w1w2_eta_cstar (cstar,eta) )},
                 c(cstarBegin, cstarEnd))$root)
}
#=====================================================================
#Tail probability of the test statistic(surrogate endpoints are used)-
#for two-stage winner design with two experimental treatments and an
#active control nu12=nu1-nu2 is the unknown treatment effect difference
#between the two experimental treatments at the interim look
#delta1 and delta2 are the target treatment effect at the final analysis
#Need to find the cstar using the function "findCutOffw12_eta_cstar" with
#given eta
#=====================================================================
gamma_w1w2 <- function(cstar, n, tau, rho, nu12, delta1,delta2,sigmaX,
sigmaY,epsilon) {
  #-------------------------------------------------------------------
  n1<-n*tau
  eta <- sqrt(tau)/2*rho
  lambda <- sqrt(n1/2/sigmaX^2)*(nu12)
  sOneEta2 <- sqrt(1-eta^2)
  k0 <- lambda / sOneEta2
  k <- eta / sOneEta2
  w1<-sqrt(n/2/sigmaY^2)*delta1
  w2<-sqrt(n/2/sigmaY^2)*delta2
  const<-sqrt(n/2/sigmaY^2)*epsilon
  #-------------------------------------------------------------------
  term1 <- integrate(function(z1e) {
      d1 <- dnorm(z1e)
      d3 <- pnorm(k0 + k*z1e)
      rvTerm1Function <- d1*d3
      return(rvTerm1Function)
    }
```

```
    , -Inf, cstar-w1-const)$value
  #-------------------------------------------------------------------------
  term2 <- integrate(function(z1e) {
      d1 <- dnorm(z1e)
      d3 <- pnorm(-k0 + k*z1e)
      rvTerm2Function <- d1*d3
      return(rvTerm2Function)
    }, -Inf, cstar-w2-const)$value
  #-------------------------------------------------------------------------
  return(1-term1-term2)
  #-------------------------------------------------------------------------
}
#============================================================================
#Sample size estimation(surrogate endpoints are used)-
#for two-stage winner design with two experimental treatments and an
#active control .nu12=nu1-nu2 is the unknown treatment effect difference
#between the two experimental treatments at the interim look
#delta1 and delta2 are the target treatment effect at the final analysis
#cstar need to determined using the function "findCutOffw12_eta_cstar"
#============================================================================
find_sample_size_two_arm <- function(nBegin, nEnd, power,cstar, tau, rho,
nu12,delta1,delta2,sigmaX,sigmaY,epsilon) {
  return(uniroot(function(n) { return(power - gamma_w1w2(cstar, n, tau,
  rho, nu12,delta1,delta2,sigmaX,sigmaY,epsilon))},
                 c(nBegin, nEnd))$root)
}
#============================================================================
#Find the Critical-value for non-inferiority test (surrogate endpoints
#are used)-
#for two-stage winner design with three experimental treatments and an
#active control
#============================================================================
library(mvtnorm)
PP <- function(b, lambda12 ,lambda13,w1, w2, w3, eta) {
```

```
  #-----------------------------------------------------------------------
sigma2 <- matrix(c(1-eta^2,.5-eta^2,.5-eta^2,1-eta^2),nrow=2)
cc <- 0*c(1,1)

P1<-function(z) {
   sapply(z, function(x) { dnorm(x)*pmvnorm(lower=-Inf,
   upper=c(lambda12+eta*(x),lambda13+eta*(x)),
                          mean=cc, sigma=sigma2 )})}
P2<-function(z) {
  sapply(z, function(x) { dnorm(x)*pmvnorm(lower=-Inf,
  upper=c(-lambda12+eta*(x),lambda13-lambda12+eta*(x)),
                          mean=cc, sigma=sigma2 )})}
P3<-function(z) {
   sapply(z, function(x) { dnorm(x)*pmvnorm(lower=-Inf,
   upper=c(-lambda13+eta*(x),lambda12-lambda13+eta*(x)),
                          mean=cc, sigma=sigma2 )})}

Int.P1 <- integrate(P1, -Inf, b-w1)$value
Int.P2 <- integrate(P2, -Inf, b-w2)$value
Int.P3 <- integrate(P3, -Inf, b-w3)$value

  return(1-Int.P1-Int.P2-Int.P3)
  #-----------------------------------------------------------------
}

#=======================================================================
#find the critical value c, when X NEQ Y
#=======================================================================
findCutOff <- function(rBegin, rEnd, alpha,eta, lambda12=0,
lambda13=0, w1=0, w2=0, w3=0 ) {
  return(uniroot(function(b) { return(alpha - PP(b,lambda12,
  lambda13,w1,w2,w3,eta) )},
                 c(rBegin, rEnd))$root)
}
```

```
# ====================================================================
#Power and Sample size computation
# ====================================================================
Power_X_NE_Y <- function(b, nu12 ,nu13,delta1, delta2, delta3,
tau,n,sigmaY,sigmaX,rho) {
  #--------------------------------------------------------------------
eta<-(sqrt(tau)/2)*rho
n1<-n*tau
w1<-sqrt(n/2/sigmaY^2)*delta1
w2<-sqrt(n/2/sigmaY^2)*delta2
w3<-sqrt(n/2/sigmaY^2)*delta3

lambda12<-sqrt(n1/2/sigmaX^2)*nu12
lambda13<-sqrt(n1/2/sigmaX^2)*nu13

sigma2 <- matrix(c(1-eta^2,.5-eta^2,.5-eta^2,1-eta^2),nrow=2)
cc <- 0*c(1,1)

P1<-function(z) {
   sapply(z, function(x) { dnorm(x)*pmvnorm(lower=-Inf,
   upper=c(lambda12+eta*(x),lambda13+eta*(x)),
                         mean=cc, sigma=sigma2 )})}
P2<-function(z) {
  sapply(z, function(x) { dnorm(x)*pmvnorm(lower=-Inf,
  upper=c(-lambda12+eta*(x),lambda13-lambda12+eta*(x)),
                         mean=cc, sigma=sigma2 )})}
P3<-function(z) {
   sapply(z, function(x) { dnorm(x)*pmvnorm(lower=-Inf,
   upper=c(-lambda13+eta*(x),lambda12-lambda13+eta*(x)),
                         mean=cc, sigma=sigma2 )})}

Int.P1 <- integrate(P1, -Inf, b-w1)$value
Int.P2 <- integrate(P2, -Inf, b-w2)$value
Int.P3 <- integrate(P3, -Inf, b-w3)$value
```

```
#cat("Int.P1=", Int.P1, "\n")
#cat("Int.P2=", Int.P2, "\n")
#cat("Int.P3=", Int.P3, "\n")

rv<-1-Int.P1-Int.P2-Int.P3
#cat("rv=", rv, "\n")

return(rv)
  #-------------------------------------------------------------------
}

# ====================================================================
#find the sample size
# ====================================================================
find_sample_size <- function(nBegin, nEnd, power,b, nu12 ,nu13,delta1,
delta2, delta3, tau,sigmaY,sigmaX,rho) {
  return(uniroot(function(n) { return(power - Power_X_NE_Y(b, nu12 ,
  nu13,delta1, delta2, delta3, tau,n,sigmaY,sigmaX,rho))},
                 c(nBegin, nEnd))$root)
}
# ====================================================================
#Find the Critical-value for non-inferiority test (Same endpoints at
#interim and final analyses)
#for two-stage winner design with two experimental treatments and an active
#control
# ====================================================================
gamma_w1w2_tau_cstar <- function(cstar,tau) {
  #-------------------------------------------------------------------------
  eta<-sqrt(tau)/2
  sOneEta2 <- sqrt(1-eta^2)
  k<-eta/sOneEta2
  #-------------------------------------------------------------------------
  term1 <- integrate(function(z1e) {
```

```
      d1 <- dnorm(z1e)
      d3 <- pnorm(k*z1e)
      rvTerm1Function <- d1*d3
      return(rvTerm1Function)
    }
    , -Inf, cstar)$value
  #-------------------------------------------------------------------------
  return(1-2*term1)
  #-------------------------------------------------------------------------
}
gamma_w1w2_tau_cstar_cut<- function(cstarBegin, cstarEnd, alpha,tau) {
  return(uniroot(function(cstar) { return(alpha - gamma_w1w2_tau_cstar
  (cstar,tau) )},
                 c(cstarBegin, cstarEnd))$root)
}

#-------------------------------------------------------------------------
power_x_eq_y_two_arm <- function(cstar, n, tau, delta1,delta2,sigmaY,epsilon) {
  #-------------------------------------------------------------------------
  eta <- sqrt(tau)/2
  w1<-sqrt(n/2/sigmaY^2)*delta1
  w2<-sqrt(n/2/sigmaY^2)*delta2
  lambda <- sqrt(tau)*(w1-w2)
  sOneEta2 <- sqrt(1-eta^2)
  k0 <- lambda / sOneEta2
  k <- eta / sOneEta2
  const<-sqrt(n/2/sigmaY^2)*epsilon
  #-------------------------------------------------------------------------
  term1 <- integrate(function(z1e) {
      d1 <- dnorm(z1e)
      d3 <- pnorm(k0 + k*z1e)
      rvTerm1Function <- d1*d3
      return(rvTerm1Function)
    }
```

```
    , -Inf, cstar-w1-const)$value
  #------------------------------------------------------------------------
  term2 <- integrate(function(z1e) {
      d1 <- dnorm(z1e)
      d3 <- pnorm(-k0 + k*z1e)
      rvTerm2Function <- d1*d3
      return(rvTerm2Function)
    }, -Inf, cstar-w2-const)$value
  #------------------------------------------------------------------------
  return(1-term1-term2)
  #------------------------------------------------------------------------
}

#-----------------------------------------------------------------------------
#find sample size
#-----------------------------------------------------------------------------
find_sample_size_two_arm_x_eqY <- function(nBegin, nEnd, power,cstar, tau, delta1,
delta2,sigmaY,epsilon) {
  return(uniroot(function(n) { return(power - power_x_eq_y_two_arm(cstar, n, tau,
  delta1,delta2,sigmaY,epsilon))},
                 c(nBegin, nEnd))$root)
}
#-----------------------------------------------------------------------------
#for two-stage winner design with three experimental treatments and an active
# control.
#Find the Critical-value for non-inferiority test
#-----------------------------------------------------------------------------
library(mvtnorm)
PP_X_EQ_Y <- function(b, w1, w2, w3, tau) {
  #------------------------------------------------------------------------
  lambda12 <- sqrt(tau)*(w1-w2)
  lambda13 <- sqrt(tau)*(w1-w3)
  #------------------------------------------------------------------------
  eta <- sqrt(tau)/2
```

```
 sigma2 <- matrix(c(1-eta^2,.5-eta^2,.5-eta^2,1-eta^2),nrow=2)
cc <- 0*c(1,1)

P1<-function(z) {
   sapply(z, function(x) { dnorm(x)*pmvnorm(lower=-Inf,upper=c(lambda12
   +eta*(x),lambda13+eta*(x)),
                          mean=cc, sigma=sigma2 )})}
P2<-function(z) {
  sapply(z, function(x) { dnorm(x)*pmvnorm(lower=-Inf,upper=c(-lambda12
  +eta*(x),lambda13-lambda12+eta*(x)),
                          mean=cc, sigma=sigma2 )})}
P3<-function(z) {
   sapply(z, function(x) { dnorm(x)*pmvnorm(lower=-Inf,upper=c(-lambda13
   +eta*(x),lambda12-lambda13+eta*(x)),
                          mean=cc, sigma=sigma2 )})}

Int.P1 <- integrate(P1, -Inf, b-w1)$value
Int.P2 <- integrate(P2, -Inf, b-w2)$value
Int.P3 <- integrate(P3, -Inf, b-w3)$value

  return(1-Int.P1-Int.P2-Int.P3)
  #-------------------------------------------------------------------------------
}
#-------------------------------------------------------------------------
#-------------------------------------------------------------------------
findCutOffb <- function(bBegin, bEnd, alpha,tau, w1=0, w2=0, w3=0 ) {
  return(uniroot(function(b) { return(alpha - PP_X_EQ_Y (b,w1, w2, w3,tau) )},
                 c(bBegin, bEnd))$root)
}
# ========================================================================
#power and sample size
# ========================================================================
library(mvtnorm)
PP_X_EQ_Y_eta_sample_epsilon <- function(b, n, sigmaY, delta1,delta2,delta3,
```

```
tau,rho=1, epsilon) {
  #-------------------------------------------------------------------------
 eta<-(sqrt(tau)/2)*rho
 w1<-sqrt(n/2/sigmaY^2)*delta1
 w2<-sqrt(n/2/sigmaY^2)*delta2
 w3<-sqrt(n/2/sigmaY^2)*delta3
 lambda12 <- 2*eta*(w1-w2)
 lambda13 <- 2*eta*(w1-w3)
 const<-sqrt(n/2/sigmaY^2)*epsilon
  #-------------------------------------------------------------------------
  sigma2 <- matrix(c(1-eta^2,.5-eta^2,.5-eta^2,1-eta^2),nrow=2)
cc <- 0*c(1,1)

P1<-function(z) {
   sapply(z, function(x) { dnorm(x)*pmvnorm(lower=-Inf,upper=c(lambda12+eta*(x),
   lambda13+eta*(x)),
                          mean=cc, sigma=sigma2 )})}
P2<-function(z) {
  sapply(z, function(x) { dnorm(x)*pmvnorm(lower=-Inf,upper=c(-lambda12+eta*(x),
  lambda13-lambda12+eta*(x)),
                          mean=cc, sigma=sigma2 )})}
P3<-function(z) {
   sapply(z, function(x) { dnorm(x)*pmvnorm(lower=-Inf,upper=c(-lambda13+eta*(x),
   lambda12-lambda13+eta*(x)),
                          mean=cc, sigma=sigma2 )})}

Int.P1 <- integrate(P1, -Inf, b-w1-const)$value
Int.P2 <- integrate(P2, -Inf, b-w2-const)$value
Int.P3 <- integrate(P3, -Inf, b-w3-const)$value
#cat("term1=", Int.P1, "\n")
#cat("term2=", Int.P2, "\n")
#cat("term3=", Int.P3, "\n")
rv <- 1-Int.P1-Int.P2-Int.P3
#cat("delta1=", delta1, "\n")
```

```
#cat("delta2=", delta2, "\n")
#cat("delta3=", delta3, "\n")
#cat("rv    =", rv, "\n")
  return(rv)
  #------------------------------------------------------------------------
}



find_sample_size_x_eq_y_epsilon <- function(nBegin, nEnd, power,b, sigmaY, delta1,
delta2,delta3, tau,rho=1, epsilon) {
  return(uniroot(function(n) { return(power - PP_X_EQ_Y_eta_sample_epsilon(b, n,
  sigmaY, delta1,delta2,delta3, tau,rho=1, epsilon))},
                 c(nBegin, nEnd))$root)
}
```

# References


[1] World Medical Association Declaration of Helsinki. Recommendations guiding physicians in biomedical research involving human subject. 2008.

[2] ICH-E10. ICH-E10,guidance for industry E10 choice of control group and related issues in clinical trials. May 2001.

[3] Weichung Joe Shih. Group sequential, sample size re-estimation and two-stage adaptive design in clinical trials: A comparison. *Statistics in Medicine*, 25:933–941, 2006.

[4] Jeff Maca. Adaptive seamless designs for phase iib/iii clinical trials. 2006.

[5] George Y. H. Chia and Qing Liu. The attractiveness of the concept of a prospectively designed two-stage clinical trial. *Journal of Biopharmaceutical Statistics*, 9:537–547, 1999.

[6] K.K. Gordon Lan, Yuhwen Soo, and Zhenming Shun. Normal approximation for two-stage winnder design. *World Scientific*, pages 28–43, 2005.

[7] Zhenming Shun, K.K. Gordon Lan, and Yuhwen Soo. Interim treatment selection using the normal approximation approach in clinical trials. *Statistics in Medicine*, 27:597–618, 2008.

[8] Genentech USA. Lucentis®(ranibizumab injection) prescribing information. 2008.

[9] David M. Brown, Peter K. Kaiser, Mark Michels, Gisele Soubrane, Jeffrey S. Heier, Robert Y. Kim, Judy P. Sy, and Susan Schneider for the ANCHOR Study Group. Ranibizumab versus verteporfin for neovascular age-related macular degeneration. *New England Journal of Medicine*, 355:1432–1444, 2006.

[10] Philip J. Rosenfeld, David M. Brown, Jeffrey S. Heier, David S. Boyer, Peter K. Kaiser, Carol Y. Chung, and Robert Y. Kim for the MARINA Study Group. Ranibizumab for neovascular age-related macular degeneration. *New England Journal of Medicine*, 355:1419–1431, 2006.

[11] Carl D. Regilloa, David M. Brownb, Prema Abrahamc, Huibin Yued, Tsontcho Ianchulevd, Susan Schneiderd, Naveed Shamsd, and The PIER Study Group. Randomized, double-masked, sham-controlled trial of ranibizumab for neovascular age-related macular degeneration: Pier study year 1. *American Journal of Ophthalmology*, 145:239–248, 2008.

[12] Peter A. Campochiaro, Jeffrey S. Heier, and Leonard Feiner et al;for the BRAVO Investigators. Ranibizumab for macular edema following branch retinal vein occlusion: six-month primary end point results of a phase iii study. *Ophthalmology*, 117:1102–1112.e1, June 2010.

[13] William C. Blackwelder. "proving the null hypothesis" in clinical trials. *Controlled clinical trials*, 3:345–353, 1982.

[14] CPMP. Points to consider on switching between superiority and non-inferiority. *http://www.tga.gov.au/docs/pdf/euguide/ewp/048299en.pdf*, 2000.

[15] H. M. James Hung, Sue-Jane Wang, Yi Tsong, John Lawrence, and Robert T. O'Neil. Some fundamental issues with non-inferiority testing in active controlled trials. *Statistics in Medicine*, 22:213–225, 2003.

[16] Steven M. Snapinn and Qi Jiang. Preservation of effect and the regulatory approval of new treatments on the basis of non-inferiority trials. *Statistics in Medicine*, 27:382–391, 2008.

[17] FDA. Guidance for industry non-inferiority clinical trials-draft guideline. March 2010.

[18] CHMP. Guideline on the choice of the non-inferiority margin. *http://www.tga.gov.au/docs/pdf/euguide/ewp/215899en.pdf*, 2005.

[19] H. M. James Hung, Sue-Jane Wang, and Robert O'Neill. Challenges and regulatory experiences with non-inferiority trial design without placebo arm. *Biometrical Journal*, 51:324–334, 2009.

[20] H. M. James Hung, Sue-Jane Wang, and Robert O'Neill. Challenges and regulatory experiences with non-inferiority trial design without placebo arm. *Biometrical Journal*, 51:324–334, 2009.

[21] Steven M. Snapinn. Alternatives for discounting in the analysis of noninferiority trials. *Journal of Biopharmaceutical statistics*, 14:263–273, 2004.

[22] Eric B. Holmgren. Establishing equivalence by showing that a specified precentage of the effect of the active control over placebo is maintained. *Journal of Biopharmaceutical Statistics*, 9:651–659, 1999.

[23] Mark Rothmann, Ning Li, Gang Chen, George Y.H. Chi, Robert Temple, and Hsiao-Hui Tsou. Design and analysis of non-inferiority mortality trial in oncology. *Statistics in Medicine*, 22:239–264, 2003.

[24] Fless JL. The statistical basis of meta-analysis. *Statistical Methods in Medical Research*, 2:121–145, 1993.

[25] Rebecca DerSimoniana and Nan Laird. Meta-analysis in clinical trials. *Controlled Clinical Trials*, 7:177–188, 1986.

[26] CBER/FDA Memorandum. Summary of cber considerations on selected aspects of active controlled trial design and analysis for the evaluation of thrombolytics in acute mi. June 1999.

[27] Steven M. Snapinn and Qi Jiang. Controlling the type 1 error rate in non-inferiority trials. *Statistics in Medicine*, 27(2008):371–381.

[28] John Whitehead. Sample size for phase 2 and phase 3 clinical trials. *Statistics in Medicine*, 5:459–464, 1986.

[29] Peter F. Thall, Richard Simon, and Susan S. Ellenberg. Two-stage selection and testing designs for comparative clinical trials. *Biometrika*, 75:303–310, 1988.

[30] Peter F. Thall, Richard Simon, and Susan S. Ellenberg. A two-stage design for choosing among several experimental treatments and a control in clinical trials. *Biometrics*, 45:537–547, 1989.

[31] Daniel J. Schaid, Sam Wieand, and Terry M. Therneau. Optimal two-stage screening designs for survival comparisons. *Biometrika*, 77:507–513, 1990.

[32] Nigel Stallard and Susan Todd. Sequential designs for phase iii clinical trials incorporating treatment selection. *Statistics in Medicine*, 22:689–703, 2003.

[33] Susan Todd and Nigel Stallard. A new clinical trial design combining phase 2 and 3: Sequential designs with treatment selection and a change of endpoint. *Drug information Journal*, 39:109–118, 2005.

[34] AR Sampson and MW Sill. Drop-the-loser design: normal case. *Biometrical Journal*, 47:257–268, 2005.

[35] Tim Friede and Nigel Stallard. A comparison of method for adaptive treatment selection. *Biometrical Journal*, 50:767–781, 2008.

[36] Qing Liu and Gordon W. Pledger. Phase 2 and 3 combination designs to accelerate drug development. *Journal of the American Statistical Association*, 100:493–536, 2005.

[37] Franz Koenig, Werner Brannath, Frank Bretz, and Martin Posch. Adaptive dunnett tests for treatment selection. *Statistics in medicine*, 27:1612–1625, 2008.

[38] Gang Li, Yining Wang, and S. Peter Ouyang. Interim treatment selection in drug development. *Statistics Bioscience*, 1:268–288, 2009.

[39] Gang Li, Junhong Zhu, S. Peter Ouyang, John Xie, Ling Deng, and Gordon Law. Adaptive designs for interim dose selection. *Statistics in Biopharmaceutical Research*, 1(4):366–376, 2009.

# Vita

Pin-wen Wang

**EDUCATION**

1999 B.S. in Public Health, National Taiwan University, Taipei, Taiwan.

2001 M.S. in Biostatistics, University of Michigan, Ann Arbor, Ann Arbor, Michigan.

2003 M.P.H. in Epidemiology, University of California, Los Angeles, Los Angeles, California.

2013 Ph.D. in Biostatistics, University of Medicine and Dentistry of New Jersey, Piscataway, New Jersey.

**EXPERIENCE**

06/00-08/00 Biostatistician as intern, PRA International, Red Bank, NJ.

01/02-12/02 Staff Research Associate, UCLA, CA.

12/02-06/03 Staff Research Associate on Cancers, UCLA, Los Angeles, CA.

09/05-05/06 Research Assistant, UMDNJ, Department of Biostatistics, Piscataway, NJ.

01/07-01/09 Biostatistician as consultant, Genentech, South San Francisco, CA.

02/09-08/09 Statistical analyst as consultant, CV Therapeutics, Palo Alto, CA.

10/10-03/13 Biostatistician as consultant, Genentech, South San Francisco, CA.