METHODS OF EVALUATION OF PERFORMANCE OF ADAPTIVE DESIGNS ON TREATMENT EFFECT INTERVALS AND METHODS OF DESIGNING TWO-STAGE WINNER DESIGNS WITH SURVIVAL OUTCOMES

BY FANG FANG

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

Yong Lin, Ph.D.

and approved by

_____

_____

_____

_____

New Brunswick, New Jersey
October, 2010

ABSTRACT OF THE DISSERTATION

METHODS OF EVALUATION OF PERFORMANCE OF ADAPTIVE DESIGNS ON TREATMENT EFFECT INTERVALS AND METHODS OF DESIGNING TWO-STAGE WINNER DESIGNS WITH SURVIVAL OUTCOMES

BY FANG FANG

Dissertation Director: Yong Lin, Ph.D.

The accuracy of the treatment effect estimation is crucial to the success of phase 3 studies. The calculation of fixed sample size relies on the estimation of the treatment effect and cannot be changed during the trial. Oftentimes, with limited efficacy data available from early phase studies and relevant historical studies, the sample size estimation may not accurately reflect the true treatment effect. Several adaptive designs have been proposed to address this uncertainty in the sample size calculation. These adaptive designs provide the flexibility of sample size adjustment during the trial by allowing early trial stopping or sample size re-estimation at the interim look(s). The use of adaptive designs can optimize the trial performance when the treatment effect is an assumed constant value. However in practice, the treatment effect is more reasonable to be considered within an interval rather than as a point estimate. Proper selection of adaptive designs will decrease the failure rate of phase 3 clinical trials and increase the chance for new drug approval. This dissertation proposes an optimal design based on an interval using the "regret concept". A mathematical framework is developed to evaluate the adaptability of different designs. In addition, this dissertation identifies the factors that may affect the performance of adap-

tive design and derives the expected sample size for two-stage sample size re-estimation designs.

In drug development, a phase 2 trial may not be feasible due to long follow-up or lack of resources. So it may necessary to evaluate several promising regimens in the confirmatory phase 3 trial. In this case, an interim analysis is often used to drop the inferior arms and to avoid the high cost, long term trial conduction, and exposure to ineffective treatment. This approach is considered as combining the two phases into one study: phase 2 portion will be carried out by the interim analysis. When appropriate surrogate endpoints exist, such as progression free survival in oncology trials, they can be used at the interim analysis to accelerate the drug development process. The statistical frameworks are available in the literature for designs with continuous endpoints. However, it is very challenging to derive the correlation between log-rank statistics at interim and final analysis when survival endpoints are used. An asymptotic correlation of log-rank statistics is developed and the features for a two-stage design survival trial using same or different endpoint at interim analysis is explored.

Preface

This document consists of two parts:

I. Evaluation of Performance for Adaptive Design on a treatment effect interval (Chapter 1 to Chapter 7).

II. Two-stage Winner Design for Survival Trials (Chapter 8 to Chapter 12).

Discussion and future directions are in Chapter 13. The dissertation is organized as follows:

- Part I:

Chapter 1 describes the challenges for traditional fixed sample size designs, research motivations and research objectives for part I.

Chapter 2 reviews the current available adaptive designs and key concepts used in this dissertation.

Chapter 3 derives the expected sample size for two-stage sample size re-estimation designs.

Chapter 4 discusses how to identify a treatment effect interval, identified a golden standard for performance comparison, and proposed the measurements of adaptive design performance based on decision theory.

Chapter 5 evaluates the performance of group sequential designs, weighted sample size re-estimation designs, and unweighted sample size re-estimation designs when treatment effects on an interval follow a uniform distribution. Comparisons are done under different design parameters.

Chapter 6 further extends the evaluation and comparison of adaptive design performance to random treatment effects. Different beta distributions are assumed for treatment effects on a pre-specified interval.

Conclusions of Part I are presented in Chapter 7.

- Part II:

Chapter 8 provides the background of two-stage winner design.

Chapter 9 presents the statistical framework on two-stage winner design for continuous endpoints and reviewed some theoretical works that Part II of this dissertation built on.

Chapter 10 develops the asymptotic correlation of log-rank statistics in two-stage winner designs when the same or different endpoint are used at the interim analysis.

Chapter 11 explores the features of two-stage winner design for survival trials

Chapter 12 presents the conclusions for Part II.

Acknowledgments

First and foremost I would like to thank my advisor, Dr. Yong Lin, whose encouragement, guidance and support from the initial to the final level enabled me to finish this dissertation. He started to train me in his advanced theory classes and guided me through the whole Ph.D. research journey. With his good teaching, sound advice, and great efforts to explain things clearly and simply, he helped to make statistics easier for me. He spent enormous time and efforts to help me from very detailed issues to huge obstacles. His inspiration and enthusiasm was motivational for me, especially during tough times in the PhD pursuit. It will be impossible for me to accomplish the goal of this research without his guidance and insight.

It is difficult to overstate my gratitude to Dr. Weichung Joe Shih, who first brought me into the adaptive design field in his class. He provided constructive suggestions at each step of this research. What I gained from him is not only the advice on my dissertation research but also the skills to conduct more research independently in the future.

I owe enormous gratitude to Dr. K.K Gordon Lan, who provided critical ideas that widely improved the validity of my work. His insights challenged me to think broadly and critically about my research.

I also want to express my sincere thanks to Dr. Shou-En Lu, who provided valuable comments on my PhD research. Further more I want to thank Dr. Lu's guidance during the first couple of years when I entered the PhD program.

Deep appreciation goes to Dr. Guangrui Zhu who provided insightful perspectives in evaluation of the performance of adaptive designs. I also want to express my

thanks individually to Drs. Yunlin Li, Jay Yang, and Xiaosha Zhang for their great discussions and collaborations in the research of two-stage winner design.

I wish to thank Eisai Medical Research, Inc. and Bayer Healthcare Pharmaceuticals, Inc for the financial support they provided during my PhD study and research period.

Lastly, and most importantly, I wish to thank my parents, Xueqin Song and Zhenjiang Fang. They bore me, raised me, taught me, loved me, and supported me spiritually throughout my life. To them I dedicate this dissertation.

Contents

List of Figures

List of Tables

Part I

Methods of Evaluation of Performance of Adaptive Designs on Treatment Effect Intervals

Chapter 1

Introduction

1.1 Challenge for Traditional Fixed Sample Size Designs

It is estimated that the cost to take a new molecular entity from laboratory to market was greater than $800 millions in 2004 [4]. The costs of drug development keeps rising at a high rate while the new drug applications are not rising at the same rate [20]. The failure rate for phase 3 trials exceeds 50% [4]. A poorly designed phase 3 trial is one possible reason for the high failure rate. It costs both money and patient lives [36]. The calculation of sample size in fixed sample size designs relies on the estimation of treatment effect and cannot be changed during the trial. The accuracy of the treatment effect estimation is crucial to the success of phase 3 studies. With limited efficacy data available from early phase studies, and relevant historical studies, the sample size estimation may not accurately reflect the true treatment effect. This lack of knowledge leads to an estimated sample size either too small or too large. Thus, the trial may be either oversized or underpowered. The results could be either a waste of finances and patient resources or trial failure.

1.2 Available Solutions

Many adaptive designs have been proposed to address this issue [2, 5, 11, 12, 14, 15, 16, 18, 21, 22, 23, 26, 29, 30, 31, 39, 41]. A broad definition for adaptive designs given by Shih [34] is used in this dissertation. All group sequential designs and sample size re-estimation designs fall into adaptive design scope per this definition. In all adaptive designs, interim analyses are built into the traditional studies; the results at the interim look(s) are used to adjust future course of the study: early stop for futility, early stop for efficacy, or to adjust sample size, etc. [34]. The tested hypotheses could be changed based on the interim results. Interim results could also be used to select an optimal dose during the study. The overall type I error rate and adequate power or conditional power need to be maintained in all adaptive designs. Detailed reviews of all available adaptive designs are provided in the literature review section.

1.3 Research Motivations

Currently, the discussion in the literature on the adaptive designs focus on a single specified value as a representation of the unknown treatment difference. However in practice, more often what is known is a treatment effect interval [23]. The following example illustrated the existence of treatment effect interval in practice.

It is believed that if an experimental drug is added to the current standard therapy (Carboplatin plus Paclitaxel), the remission time for ovarian cancer patients after surgery will be prolonged. A phase 3 confirmatory trial is planned to compare the treatment effect of the combination therapy versus the standard therapy alone. Progression free survival

(PFS) is used as the primary efficacy endpoint. The treatment effect is estimated based on the results of phase 2 proof of concept studies for the experimental drug and published median PFS for the standard treatment. However, different PFS medians for standard therapy are found in the literature. In the Hellenic Cooperative Oncology Group (HeCOG) study [1], 121 out of 247 patients were randomized to the standard (Carboplatin plus Paclitaxel) therapy arm. At the end of study, the median PFS for standard therapy arm was 38 months. In the Gynecologic Oncology Group (GOG) study, the median PFS for 392 patients who received standard therapy was 19.4 months [27]. In another phase 3 study supported by Bristol-Meyers Squibb (BMS), the median PFS was 16 months for patients who received standard therapy [25]. Two other randomized trials indicated the median PFS was around 17.5 months [6, 28]. The standard therapy regimens including dose levels and dose frequencies were similar for these studies. There were also differences in cancer stages and tumor sizes among the patients populations enrolled into these studies. Thus, it became very difficult to find an accurate point estimate of the true median PFS for the standard therapy. After careful comparison of study designs including inclusion/exclusion criteria and treatment schedules, the median PFS for standard therapy was most likely between 15 to 20 months. Thus, based on this information and the data from early proof of concept studies on the experimental drug, it was expected that the treatment effect between the standard therapy and the combination therapy would fall into a certain interval.

Most of the previous research and designs were focused on how to maximize the study productivity when the treatment effect was estimated as a point value. For the examples above, the question arises as to how the current available adaptive designs can be

used to maximize the study productivity on a treatment effect interval? How can we use mathematical frame to evaluate the performance of different designs? What factors will affect the performance of adaptive designs? Under the same constrains, whether certain group sequential designs (e.g., with different boundaries or different sample size increments) have a similar performance as sample size re-estimation designs?

1.4 Research Objectives

1.4.1 Objectives

In Part I of this dissertation, the following objectives were set:

Objective I: To derive the expected sample size for sample size re-estimation design with a total of 2 looks (Chapter 3).

Objective II: To evaluate and compare the performance of adaptive designs including group sequential designs, weighted sample size re-estimation designs, and unweighted sample size re-estimation designs on a pre-specified treatment effect interval when treatment effects follow a uniform distribution (Chapter 4).

Objective III: To evaluate and compare the performance of adaptive designs on a pre-specified treatment effect interval when treatment effects follow a Beta distribution (Chapter 5).

1.4.2 Specific aims

The specific aims for Part I of this dissertation are described as follows:

1. To derive expected sample size for two-stage sample size re-estimation designs

2. To determine an ideal design to be used as the gold standard for the comparison of performance

3. To identify appropriate measurements of performance on the treatment effect interval

4. To compare performance of adaptive designs by using the appropriate measurement(s) identified

5. To compare designs with different sample size increments including equal spaced increments and 2-times increments

6. To estimate the performance of adaptive designs on intervals with different adaptive indexes

7. To evaluate effect of types of boundaries on the performance of adaptive designs

8. To considere treatment effect as a random variable follows a beta distribution and evaluate the performance of the adaptive designs.

Chapter 2

Literature Review

For a variety of reasons, interim analyses are planned in clinical trials. One of the most important reasons is the ethics to avoid exposing patients to unsafe or ineffective treatment [15]. Another crucial reason is to ensure that studies have adequate sample size and power by checking important protocol assumptions including treatment effects, variance, etc.

Wittes and Brittain proposed to using an internal pilot study to recalculate the required sample size during the study through re-estimation of variance based on unblinded data [30]. Gould and Shih proposed a sample size re-estimation strategy without unblinding of treatment codes [11, 12] to evaluate the variance.

There are several other approaches to re-evaluate the treatment effect (assume a known variance) at the interim looks proposed. Different group sequential designs are available to allow stopping the trial early for benefit or futility by doing interim analyses [26, 30]. Several designs are proposed to provide flexibility to adjust the sample size

through sample size re-estimation built into the group sequential designs [5, 14]. More reviews on the re-estimation of treatment effect are provided in the sections below.

2.1 Group Sequential Designs

Since repeated significance tests need to be done after each observation or matched pair of observations in the traditional sequential designs, this is often impracticable [30], Pocock proposed a group sequential test in 1977. In his design, patients were divided into several equal-sized (equal-spaced) groups. The decision to stop or to continue the trial was to be made based on repeated significance tests of accumulated data after each group of patients [30]. A group sequential design for a normal response with known variance was defined in this article. A numeric method was used to find the total sample size and the nominal alpha (Type I error rate) value for each significance test. The nominal alpha values (critical values) were the same for repeated significance tests. It was shown by simulation studies that a group sequential design can also be used for other types of response data other than a normal response.

O'Brien and Fleming proposed a similar multiple testing procedure. As in the Pocock design, the significance tests were conducted after accrual of each group of equal-sized patients based on the accumulated data. However, the critical values were different for each test. The earlier the look was, the larger the critical value was. A numeric method was used to obtain the critical values [26]. Compared to Pocock boundaries, the O'Brien-Fleming boundary makes it more difficult to stop the trial at early stage.

Haybittle [14] and Peto et al [29] suggested an approach using a fixed z-score

boundary of 3 at the interim analyses and using critical value of 1.96 at the final analysis. Because of the high critical values at the interim looks, the chance for early stopping under the null hypothesis was very limited and the inflation of type I error was very small. However, the more the number of interim looks was, the larger the inflation of type I error was. In order to maintain the overall type I error at the exact level, a modified Haybittle-Peto method can be used. Instead of using 1.96, the final critical value needs to be adjusted to ensure the overall α level.

2.2 Sample Size Re-estimation Designs

While group sequential designs are based on a fixed maximum number of patients or events, sample size re-estimation designs provide the flexibility to adjust the sample size at the interim look. A considerable amount of research was done in recent years on the basis of updated information.

Cui, Hung, and Wang [5] proposed a new adaptive design in 1999 by modifying the weights used in the traditional repeated two-sample mean test. In their article, the authors elaborated the need of implementation of a valid inferential procedure that allowed flexibility of adjusting the sample size based on the unblinded estimate of treatment effect during the study. They evaluated the impact of sample size change based on the interim estimation of treatment effect and demonstrated that the large gain in power increase at the cost of a trivial inflation in type I error rate through simulation studies.

Because of the sample size re-estimation, the two sample mean test is no longer a Brownian motion process [5]; and the total sample size becomes a random variable de-

pending on the observed treatment effect at the interim look [5]. So obviously the traditional repeated two-sample mean test can not be used in this situation. Cui, Hung, and Wang [5] proposed a new test procedure. In this method, the test statistic used to determine statistical significance at each interim look was a weighted combination of two independent Wald statistics, comprising the data before and after the sample size adaptation. This new procedure used the same critical values as those for the traditional repeated two-sample mean test. It was also demonstrated that the total type I error rate for the new procedure was equal to the original test procedure. Below, some of the details of this new procedure are summarized.

2.2.1 Traditional repeated two-sample mean test

First, let's consider using the traditional repeated two-sample mean tests to detect the mean difference in two independent normal populations with a known common variance σ^2. Suppose $\mu_1$and μ_2 are the means for normal populations x and y and let N be the total sample size for each treatment and $\Delta = \mu_1 - \mu_2$. Then we have $x_i \sim N\left(\mu_1, \sigma^2\right)$, $y_i \sim N\left(\mu_2, \sigma^2\right)$, and $\overline{x} - \overline{y} \sim N\left(\mu_1 - \mu_2, \frac{2\sigma^2}{N}\right)$, where i=1, 2, …, N. There are total K looks. Assume N_L and N_{L+J} are the sample sizes at L^{th} and $(L+J)^{th}$ interim looks and $1 \leq L < L+J \leq K$. Assume Δ_L is the observed treatment difference at interim look L and two-sample mean test statistic is T_L. Then the test statistic at $(L+J)^{th}$ look can be denoted as

$$T_{L+J} = \frac{\left(\sum_{i=1}^{N_{L+J}} x_i - \sum_{i=1}^{N_{L+J}} y_i\right) / N_{L+J}}{\sqrt{\frac{2\sigma^2}{N_{L+J}}}} = T_L \sqrt{\frac{N_L}{N_{L+J}}} + W_{L+J} \sqrt{\frac{N_{L+J} - N_L}{N_{L+J}}}$$

where $T_L = \frac{\sum_{i=1}^{N_L}(x_i - y_i)}{\sqrt{2\sigma^2 N_L}}$ and $W_{L+J} = \frac{\sum_{i=N_L+1}^{N_{L+J}}(x_i - y_i)}{\sqrt{2\sigma^2\left(N_{L+J} - N_L\right)}}$.

T_{L+J} can be decomposed into two statistics (T_L and W_{L+J}) that weighted by the amount of information available before the L^{th} interim look ($\sqrt{\frac{N_L}{N_{L+J}}}$) and the information after L^{th} interim look $\left(\sqrt{\frac{N_{L+J}-N_L}{N_{L+J}}}\right)$.

2.2.2 New group sequential test procedure proposed by Cui-Hung-Wang (CHW)

CHW proposed a new procedure for study with sample size re-estimation built in at L^{th} interim look. Let M_{L+J} be the new sample size at$(L+J)^{th}$ interim look. Then the test statistic after sample size re-estimation for the new procedure at $(L+J)^{th}$ interim analysis is

$$U_{L+J} = T_L\sqrt{\frac{N_L}{N_{L+J}}} + W^*_{L+J}\sqrt{\frac{N_{L+J}-N_L}{N_{L+J}}}$$

where $W^*_{L+J} = \frac{\sum_{i=N_L+1}^{M_{L+J}}(x_i-y_i)}{\sqrt{2\sigma^2\left(M_{L+J}-N_L\right)}}$.

Compared to the traditional repeated test procedure, W^*_{L+J} is used in the test statistic for the $(L+J)^{th}$ interim test. W^*_{L+J} is a random variable based on Δ_L rather than a fixed amount of information [16]. The weights for the two decomposed statistics, regardless the new sample size after the L^{th} interim analysis, are the same as the traditional tests. That is the weights in the new procedure are still based on the original sample size before the sample size re-estimation.

Both $(T_1, T_2, \ldots, T_L, \ldots, T_K)$ and $(T_1, T_2, \ldots, U_L, \ldots, U_K)$ follow the same multivariate normal distribution under the null hypothesis. Thus the total type I error rate for the new test statistics will be the same as of the original test procedure without sample size increase.

Proof:

$$\begin{aligned} E\left(U_{L+J}T_L\right) &= E\left(T_L\sqrt{\frac{N_L}{N_{L+J}}} + W^*_{L+J}\sqrt{\frac{N_{L+J}-N_L}{N_{L+J}}}|T_L\right) \\ &= E\left(T_L\sqrt{\frac{N_L}{N_{L+J}}} + \frac{\sum_{i=N_{L+1}}^{M_{L+J}}(x_i-y_i)}{\sqrt{2\left(M_{L+J}-N_L\right)}}\sqrt{\frac{N_{L+J}-N_L}{N_{L+J}}}|T_L\right) \\ &= T_L\sqrt{\frac{N_L}{N_{L+J}}} + \sqrt{\frac{N_{L+J}-N_L}{N_{L+J}}}E\left(\frac{\sum_{i=N_{L+1}}^{M_{L+J}}(x_i-y_i)}{\sqrt{2\left(M_{L+J}-N_L\right)}}\right). \end{aligned}$$

Under H_0, $E\left(\frac{\sum_{i=N_{L+1}}^{M_{L+J}}(x_i-y_i)}{\sqrt{2\left(M_{L+J}-N_L\right)}}\right) = 0$, we have

$$E\left(U_{L+J}|T_L\right) = E\left(T_{L+J}|T_L\right) = T_L\sqrt{\frac{N_L}{N_{L+J}}},$$

$$\begin{aligned} Var\left(U_{L+J}T_L\right) &= Var\left(T_L\sqrt{\frac{N_L}{N_{L+J}}} + W^*_{L+J}\sqrt{\frac{N_{L+J}-N_L}{N_{L+J}}}|T_L\right) \\ &= Var\left(\frac{\sum_{i=N_{L+1}}^{M_{L+J}}(x_i-y_i)}{\sqrt{2\left(M_{L+J}-N_L\right)}}\sqrt{\frac{N_{L+J}-N_L}{N_{L+J}}}|T_L\right) \\ &= \sqrt{\frac{N_{L+J}-N_L}{N_{L+J}}}Var\left(\frac{\sum_{i=N_{L+1}}^{M_{L+J}}(x_i-y_i)}{\sqrt{2\left(M_{L+J}-N_L\right)}}\right) \\ &= \frac{N_{L+J}-N_L}{N_{L+J}} \\ &= Var\left(T_{L+J}|T_L\right), \end{aligned}$$

$$cov\left(U_{L+I},U_{L+J}\right) = \sqrt{\frac{N_{L+I}-N_L}{N_{L+I}}}\sqrt{\frac{N_{L+J}-N_L}{N_{L+J}}}cov\left(W^*_{L+I},W^*_{L+J}\right),$$

and

$$\begin{aligned} cov\left(W^*_{L+I},W^*_{L+J}\right) &= cov\left(\frac{\sum_{i=N_{L+1}}^{M_{L+I}}(x_i-y_i)}{\sqrt{2\left(M_{L+I}-N_L\right)}},\frac{\sum_{i=N_{L+1}}^{M_{L+J}}(x_i-y_i)}{\sqrt{2\left(M_{L+J}-N_L\right)}}\right) \\ &= cov\left(\frac{S_{L+I}-S_L}{\sqrt{2\left(M_{L+I}-N_L\right)}},\frac{S_{L+J}-S_L}{\sqrt{2\left(M_{L+J}-N_L\right)}}\right) \end{aligned}$$

$$
\begin{aligned}
&= \frac{var\left(S_{L+\min(I,J)} - S_L\right)}{2\sqrt{M_{L+I} - N_L}\sqrt{M_{L+J} - N_L}} \\
&= \frac{2\left(M_{L+\min(I,J)} - N_L\right)}{2\sqrt{M_{L+\min(I,J)} - N_L}\sqrt{M_{L+\max(I,J)} - N_L}} \\
&= \sqrt{\frac{M_{L+\min(I,J)} - N_L}{M_{L+\max(I,J)} - N_L}} \\
&= \sqrt{\frac{\left(b\left(M_{L+\min(I,J)} - N_L\right) + N_L\right) - N_L}{\left(b\left(M_{L+\max(I,J)} - N_L\right) + N_L\right) - N_L}} \\
&= \sqrt{\frac{N_{L+\min(I,J)} - N_L}{N_{L+\max(I,J)} - N_L}}.
\end{aligned}
$$

So we have

$$
cov\left(U_{L+I}, U_{L+J}|T_L\right) = \sqrt{\frac{N_{L+I} - N_L}{N_{L+I}}}\sqrt{\frac{N_{L+J} - N_L}{N_{L+J}}}\sqrt{\frac{N_{L+\min(I,J)} - N_L}{N_{L+\max(I,J)} - N_L}} = cov\left(T_{L+I}, T_{L+J}|T_L\right)
$$

and $(T_1, T_2, \ldots, T_L, \ldots, T_K)$ and $(T_1, T_2, \ldots, U_L, \ldots, U_K)$ follow the same multivariate normal distribution under the null hypothesis.

2.2.3 Comparison of three two-stage designs

Shih [34] summarized the commonality of adaptive designs (including both group sequential designs and sample size re-estimation designs). All adaptive designs involve interim analyses and need to maintain the overall type I error probability. Group sequential designs were developed for early stopping to reject or accept H_0; the maximum information was fixed and required at the planning stage; the primary efficacy endpoints always involve mortality or irreversible morbidity [34]. Sample size re-estimation designs were usually designed to reinforce power by checking assumptions and re-estimating sample size (blinded or unblinded analysis, or internal pilot study); the total information at the

planning stage can be expanded; and the primary efficacy endpoints were often non-life-threatening measures [34]. It is not like the group sequential designs in which interim analysis can be done at very early stage, the sample size at interim look for sample size re-estimation design should not be too small [34]. This is because the primary purpose for sample size re-estimation is to reinforce the study power based on the information gathered at the interim look and enough amount of information is necessary to ensure the quality of sample size re-estimation. In this article, the author also compared a couple of two-stage adaptive designs -CHW [5] and LR-AD [21, 22] with the traditional two-stage group sequential design (GS).

Because of the simplicity of two-stage designs and its frequent use in practice, Shih [34] chose three two stage designs to illustrate the characteristics of different adaptive designs. Let $n_1 + m$ be the fixed maximum number of samples, n_1 and m denote the sample sizes at stage 1 and stage 2 respectively; and n_2 denote the new sample size for stage 2. Below are shown the test statistics for the group sequential design, CHW sample size re-estimation design, and likelihood ratio design, where $Z_2(n_1)$ indicates that Z_2 is dependent of the sample size n_1 at stage 1 and $Z_2(m)$ indicates that Z_2 is independent of the sample size at stage 1.

$$
\begin{aligned}
Z_{GS} &= \sqrt{\frac{n_1}{n_1+m}}Z_1 + \sqrt{\frac{m}{n_1+m}}Z_2\,(m)\,,\\
Z_{CHW} &= \sqrt{\frac{n_1}{n_1+m}}Z_1 + \sqrt{\frac{m}{n_1+m}}Z_2\,(n_1)\,,\\
Z_{LR} &= \sqrt{\frac{n_1}{n_1+n_2}}Z_1 + \sqrt{\frac{n}{n_1+n_2}}Z_2\,(n_1)\,.
\end{aligned}
$$

The advantage of CHW and LR-AD designs is their flexibility. Sample size can be re-evaluated after interim analysis in LR-AD and CHW designs while GS design has a

fixed maximum sample size. Through sample size re-estimation, power can be reinforced in LR-AD and CHW designs [5]. In Figure 2.1 , the difference among these three two-stage designs was further illustrated.

It can be seen that even though the sample size re-estimation will be done for the CHW design, the final statistic will be rescaled according to the original maximum information and the portions of each stage. So in terms of critical values, CHW design and group sequential design have the same critical values at the interim analysis and the final analysis. As long as the maximum information for GS or the initial maximum information for CHW is determined, the critical values are determined and will not be affected by the interim analysis. This is not the case for the LR-AD design, its' critical values depend on the conditional power and the type I error left for the final look.

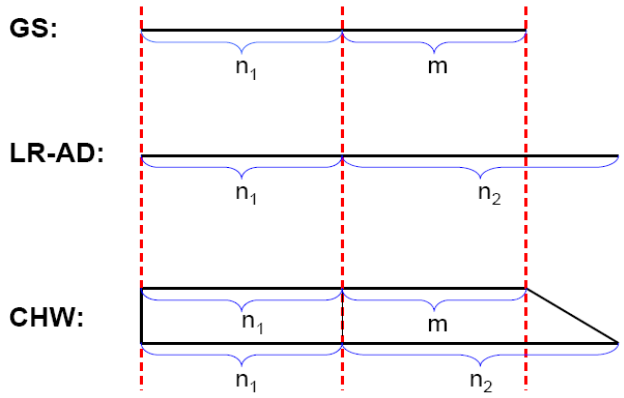

Figure 2.1: Comparison of Two-stage Designs

2.2.4 Evaluation of Adaptability – Comparison of Adaptive Designs

Liu *et al*. [23] talked about the process of selecting a flexible sample size design based on a proposed performance score, that measured the overall adaptive performance when the treatment effect was in a pre-determined interval $[\delta_L, \delta_U]$. The authors considered a one-sided hypothesis test with a significance level of α. The variance is known and is equal to 1. The true treatment difference is δ and δ_0 is the pre-specified treatment difference when the trial is designed. The ideal design is to achieve $100\,(1-\beta)$ per cent power and have the ideal fixed sample size which is unachievable because δ is never known in practice. Fixed sample size can be calculated as

$$N_{1-\beta}\left(\delta_0\right) = \frac{\left(Z_{1-\beta} + Z_{1-\alpha}\right)^2}{\delta_0^2}.$$

The following definitions are used in the paper:

Sample size ratio is defined as:

$$SR\left(N|\delta,\beta\right) = \frac{Study\ sample\ size}{Ideal\ Sample\ size}$$

$$
\begin{aligned}
&= \frac{N_{1-\beta}\left(\delta_0\right)}{N_{1-\beta}\left(\delta\right)} \\
&= \frac{\frac{\left(Z_{1-\beta}+Z_{1-\alpha}\right)^2}{\delta_0^2}}{\frac{\left(Z_{1-\beta}+Z_{1-\alpha}\right)^2}{\delta^2}} \\
&= \frac{\delta^2}{\delta_0^2}.
\end{aligned}
$$

Measure the adaptive performance of a flexible sample size design using the relative oversize (ROS) and underpower (RUP) functions:

$$
\begin{aligned}
ROS\left(\delta|f_s,\beta\right) &= E\left(SR\left(N|\delta,\beta\right)-1\right)_+ / \left(f_s-1\right)\times 100\%, \\
RUP\left(\delta|f_p,\beta\right) &= \left[N_{1-\beta}-N_{pow}\right]_+ / \left[N_{1-\beta}-N_{\left(1-f_p\right)\times(1-\beta)}\right]\times 100\%,
\end{aligned}
$$

where 'pow' denotes the power value of a given flexible sample size design at the true treatment difference δ, $N_{pow} = (Z_{pow}+Z_{1-\alpha})^2/\delta^2$ is the corresponding sample size, and $f_s > 1$ and $f_p < 1$ are scaling factors. These two scaling factors can be adjusted depending on the considerations for underpower and oversize. If underpower is viewed as a more serious matter, then a small value of f_p may be used to penalize even a small reduction in power from the target value. On the other hand, a small f_s may be used to penalize even a small amount of oversize when oversize is the major concern[23].

Total performance function:

$$
R\left(\delta|f_p,f_s,\beta\right) = RUP\left(\delta|f_p,\beta\right) + ROS\left(\delta|f_s,\beta\right)
$$

Average performance score (APS):

$$
APS\left(f_p,f_s,\beta\right) = \int_{\delta_L}^{\delta_U} R\left(\delta|f_p,f_s,\beta\right) w\left(\delta\right) d\delta
$$

where $w(\delta)$ is a weight function. If it is believed that the true treatment effect is equally likely over the given interval, a uniform weight may be used. APS value is the smaller the better. A smaller value of APS indicates that design is closer to the optimal performance on the considered interval [23].

Simulation studies were conducted to demonstrate that APS provides a useful tool for the evaluation of study designs and for the determination of a better sample size strategy through the minimization of APS across designs. In the simulation studies, equal-spaced GS designs with total 4 looks by using O'Brien Fleming or Pocock boundaries were used. The best design with the smallest APS was selected among GS designs in each design category with maximum sample size varying from 40 to 1000 with an increment of 40 [23].

It was shown that both the GS designs with Pocock boundaries and O'Brien Fleming boundaries had much smaller APS compared to fixed sample size designs. GS design with O'Brien Fleming boundaries had a slightly higher APS then the designs with Pocock boundaries.

The APS was further improved by allowing sample size re-estimation during the study. However as the authors pointed out in their conclusion, that it was largely due to the fact that the maximum sample size for the GS designs cannot be changed while sample size re-estimation designs can increase the sample size based on the sample size re-estimation. It also needs to be considered whether it is a fair comparison when different maximum sample sizes are allowed for different designs. This maximum sample size could either be the pre-determined maximum sample size or via sample size re-estimation.

2.3 Key Concepts Used for Adaptive Design Research

2.3.1 Study Level Type I Error Inflation

For adaptive designs, the type I error (overall significance level) is the probability of at least one significant difference when the null hypothesis is true. Based on Armitage et al., type I error increases when repeated significance tests are done [2]. The total probability of rejecting H_0 increases when the number of repeated tests (type I error is 0.05 for each test) grows. Examples are provided in the Table 2.1 [15] below.

Table 2.1: Inflation of Type I error

Number of Tests	Overall Null Probability of Rejecting H_0*
1	0.05
2	0.08
3	0.11
4	0.13
5	0.14
10	0.19
20	0.25
50	0.32

* Type I error for each test is 0.05.

To overcome this problem, more stringent nominal significance levels for each repeated test at interim analysis need to be chosen to maintain the overall significance level at a pre-specified levels.

2.3.2 Type of Boundaries for Group Sequential Designs

In general, there are three types of boundaries. Pocock boundaries use the same critical value at each look. Similar to Pocock boundaries, Haybittle-Peto type boundaries

for the first k-1 interim looks are the same. The boundary for the final test is similar to the fixed sample size design and ensures the total alpha at a pre-specified level. O'Brien-Fleming boundaries are more stringent at early looks compared to the other two types of boundaries. Thus, it is more difficult to stop the trial at early stage by using a O'Brien-Fleming boundary. Below is an example for a clinical trial with a total of 6 looks, a two-sided type I error rate at 0.05, and the critical value for the first 5 looks for the Haybittle-Peto boundaries set at α_0=0.005 level. Boundaries at each interim look are provided in the Table 2.2 and Figure 2.2 below.

Table 2.2: Boundaries from Exact Method

		Type of Boundary		
Look	Information Fraction	O'Brien_Fleming	Pocock	Modified Haybittle-Peto ($\alpha_0 = 0.005$)
1	0.1667	5.0283	2.4532	2.8070
2	0.3333	3.5555	2.4532	2.8070
3	0.5000	2.9031	2.4532	2.8070
4	0.6667	2.5142	2.4532	2.8070
5	0.8333	2.2487	2.4532	2.8070
6	1	2.0528	2.4532	2.0400

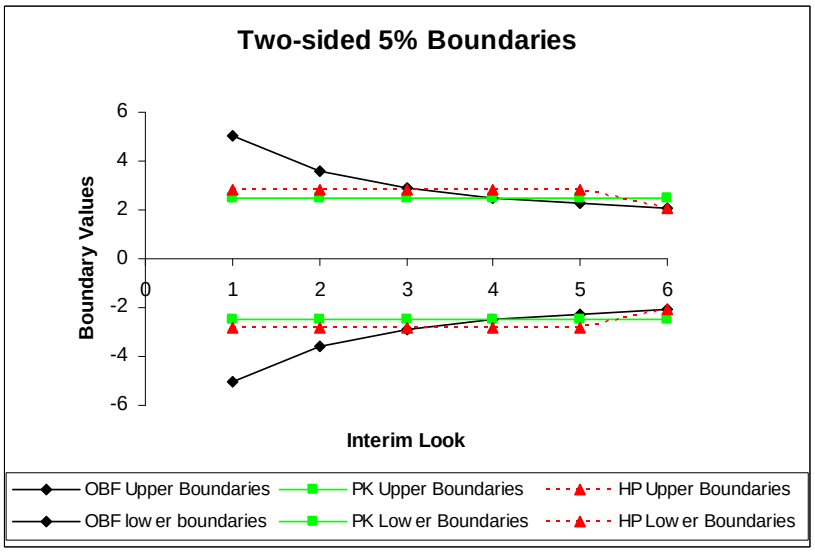

Figure 2.2: Different Types of Boundaries

Boundaries from Exact Method Different type of boundaries can be calculated by using numeric method. Details for boundary calculation are provided in section 2.3.5. Let $t_1, t_2, ..., t_K$ denote information fraction at each look, δ denote the treatment effect, and $Z_{t_1}, Z_{t_2}, ..., Z_{t_K}$ denote the test statistic at each look. Then,

1. Pocock Boundaries satisfy

$$P\left(Z_{t_1} > c, \text{ or } Z_{t_2} > c, \ldots \text{ or } Z_{t_K} > c | \delta = 0\right) = \alpha.$$

2. O'Brien-Fleming Boundaries satisfy

$$P\left(Z_{t_1}\sqrt{t_1} > c, \text{ or } Z_{t_2}\sqrt{t_2} > c, \ldots \text{ or } Z_{t_K}\sqrt{t_K} > c | \delta = 0\right) = \alpha.$$

3. Haybittle-Peto Boundaries satisfy

$$P\left(Z_{t_1} > c_0, \text{ or } Z_{t_2} > c_0, \ldots \text{ or } Z_{t_K} > c_{\alpha-\alpha(K-1)} | \delta = 0\right) = \alpha,$$

where $c_0 = \Phi^{-1}(1-\alpha_0)$, α_0 is predetermined and $\alpha_0(K-1)$ is the cumulative alpha spending in the first $K-1$ looks.

Boundaries from Alpha Spending Function

All types of boundaries discussed above are discrete sequential boundaries. The total number of looks needs to be specified in advance. Lan and DeMets [18] proposed a flexible way to construct discrete sequential boundaries by using a function $\alpha^*(t)$ where t is the decision time (information fraction). The boundary at a decision time will be determined by $\alpha^*(t)$ and will not be affected by the future time and total number of looks. The proposed function for Pocock boundaries is $\alpha^*(t) = \alpha \ln[1+(e-1)t]$ and the proposed function for O'Brien-Fleming boundaries is $\alpha^*(t) = 2\left(1-\Phi\left(\frac{Z_{\alpha/2}}{\sqrt{t}}\right)\right)$. Boundaries for the above example by using alpha spending functions are provided in the table below. They are very close to the boundaries from the exact method.

Table 2.3: Boundaries from alpha Spendng Approach

		Type of Boundary	
Look	Information Fraction	O'Brien-Fleming	Pocock
1	0.1667	5.3667	2.4951
2	0.3333	3.7103	2.4769
3	0.5000	2.9697	2.4550
4	0.6667	2.5387	2.4373
5	0.8333	2.2522	2.4233
6	1	2.0448	2.4121

2.3.3 Bivariate Normal Distributions for Alpha Adjustment Calculations

Bivariate normal distribution plays an important role in calculating boundaries and conditional power. Derivations for mean and variance of bivariate normal distribution under adaptive design setting are provided below. Suppose the treatment effect variables follow normal distributions with equal variance, i.e. $x_i \sim N(\mu_1, 1)$, $y_i \sim N(\mu_2, 1)$, where $i = 1, 2, 3, ..., n$. Let $\Delta = \mu_1 - \mu_2$, then $d_i = x_i - y_i \sim N(\Delta, 2)$, $S_n = \sum_{i=1}^{n} d_i \sim N(n\Delta, 2n)$, and $Z_n = \frac{S_n}{\sqrt{2n}} \sim N\left(\sqrt{\frac{n}{2}}\Delta, 1\right)$. Let N_{max} denote the maximum sample size, n_i denotes the sample size at interim look i, and t_i denotes the information fraction at interim look i and $t_i = n_i / N_{max}$ and $n_i = N_{max} * t_i$.

The test statistics for any two interim looks will follow a bivariate normal distribution. The test statistics when interim analysis are done after accrual n_1 and n_2 $(> n_1)$ patients are

$$Z_1 = \frac{S_{n_1}}{\sqrt{2n_1}} \sim N\left(\sqrt{\frac{n_1}{2}}\Delta, 1\right)$$

and

$$Z_2 = \frac{S_{n_2}}{\sqrt{2n_2}} \sim N\left(\sqrt{\frac{n_2}{2}}\Delta, 1\right).$$

The covariance between Z_1 and Z_2 is

$$\begin{aligned} cov(Z_1, Z_2) &= cov\left(\frac{S_{n_1}}{\sqrt{2Nt_1}}, \frac{S_{n_2}}{\sqrt{2Nt_2}}\right) \\ &= cov\left(\frac{S_{n_1}}{\sqrt{2Nt_1}}, \frac{(S_{n_2} - S_{n_1}) - S_{n_1}}{\sqrt{2Nt_2}}\right) \\ &= cov\left(\frac{S_{n_1}}{\sqrt{2Nt_1}}, \frac{S_{n_1}}{\sqrt{2Nt_2}}\right) \\ &= \frac{1}{\sqrt{2Nt_1}} \frac{1}{\sqrt{2Nt_2}} var(S_{n_1}) \end{aligned}$$

$$= \frac{1}{\sqrt{2Nt_1}} \frac{1}{\sqrt{2Nt_2}} 2n_1$$
$$= \sqrt{\frac{t_1}{t_2}}.$$

Thus

$$\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \sqrt{\frac{n_1}{2}}\Delta \\ \sqrt{\frac{n_2}{2}}\Delta \end{pmatrix}, \begin{pmatrix} 1 & \sqrt{\frac{t_1}{t_2}} \\ \sqrt{\frac{t_1}{t_2}} & 1 \end{pmatrix} \right).$$

2.3.4 Conditional Power at Interim

Definition of conditional power

Conditional power is the conditional probability of a significant result at the end of the trial given the data observed thus far [31]. Conditional power is a very useful tool to define the early stop for futility rule in clinical trials. It also can be used to calculate the new sample size in the sample size re-estimation design. It is a function of treatment effect Δ. Usually conditional power is calculated under different assumptions about the trend of the future data. The trend can be assumed under the current data, under the null hypothesis, or under the alternative hypothesis as shown in Figure 2.3.

Calculation of conditional power

Conditional power can be calculated as follow:

As we know that when $\binom{x_1}{x_2} \sim N\left(\binom{\mu_1}{\mu_2}, \begin{pmatrix} \sum_{11} & \sum_{12} \\ \sum_{21} & \sum_{22} \end{pmatrix} \right)$, the distribution for $x_2|x_1$will be

$$x_2|x_1 \sim N\left(\mu_2 + \sum_{21}\sum_{11}^{-1}(x_1 - \mu_1), \sum_{22} - \sum_{21}\sum_{11}\sum_{12}\right).$$

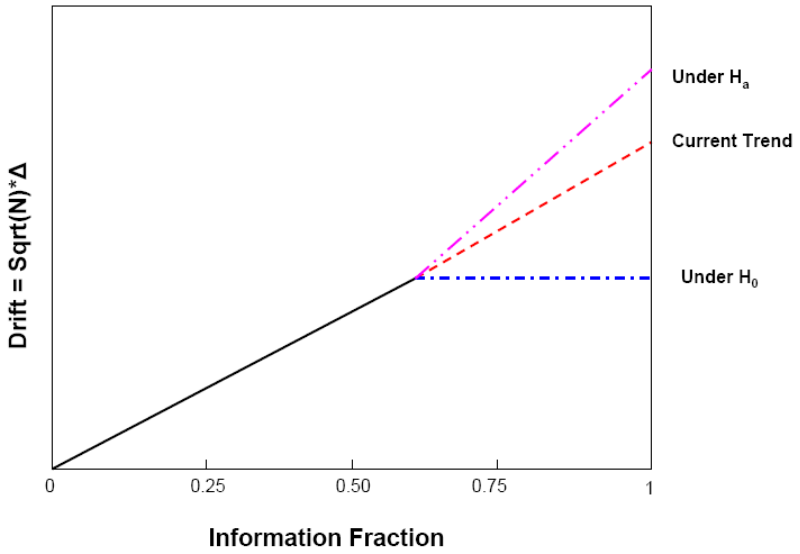

Figure 2.3: Different trends for calculation of conditional power

In the adaptive design setting, $x_1$is the Z_1 statistic based on the available data thus far and Z_2 is the statistic for a future analysis. Then

$$(Z_2|Z_1) \sim N\left(Z_1\sqrt{\frac{t_1}{t_2}} + \frac{\sqrt{N}\,(t_2 - t_1)}{2\sqrt{t_2}}\Delta, 1 - \frac{t_1}{t_2}\right),$$

where N is the sample size and Δ is treatment difference under the trend assumption. The conditional power can be derived by plugging into the above mean and standard error into the formula below.

$$\begin{aligned} C_P &= P_\Delta\left(Z_2 > c_2|Z_1\right) \\ &= P_\Delta\left(\frac{Z_2 - E\left(Z_2|Z_1\right)}{\sqrt{var\left(Z_2|Z_1\right)}} > \frac{c_2 - E\left(Z_2|Z_1\right)}{\sqrt{var\left(Z_2|Z_1\right)}}\right) \\ &= 1 - \Phi\left(\frac{c_2 - E\left(Z_2|Z_1\right)}{\sqrt{var\left(Z_2|Z_1\right)}}\right) \\ &= 1 - \Phi\left(\frac{c_2 - Z_1\sqrt{\frac{t_1}{t_2}} - \frac{\sqrt{N}(t_2 - t_1)}{2\sqrt{t_2}}\Delta}{\sqrt{1 - \frac{t_1}{t_2}}}\right). \end{aligned}$$

As an example, the conditional power formula used in the simulation study for the sample size re-estimation in CHW design is provided below. Let $n_{\max}$ be the initial sample size, n_L be the cumulative sample size at interim look for sample size re-estimation, $m_{\max}$ be the new sample size after sample size re-estimation, and c_K be the critical value at final analysis. The conditional power for CHW design is

$$\begin{aligned} C_P &= P\left(Z_K > c_K | Z_L\right) \\ &= P\left(\sqrt{\frac{n_L}{n_{\max}}} Z_L + \sqrt{1 - \frac{n_L}{n_{\max}}} Z^* > c_K|_{Z_L}\right) \\ &= P\left(Z^* > \frac{c_K - \sqrt{\frac{n_L}{n_{\max}}} Z_L}{\sqrt{1 - \frac{n_L}{n_{\max}}}} | Z_L\right) \\ &= P\left(Z^* > c_K \sqrt{1 + \frac{n_L}{n_{\max} - n_L}} - Z_L \sqrt{\frac{n_L}{n_{\max} - n_L}} | Z_L\right), \end{aligned}$$

where Z^* is the test statistic based on $M_{max} - N_L$ future observations. Because Z^* is independent of Z_L, $Z^*|Z_L = Z^* \sim N\left(\sqrt{\frac{n_{\max} - n_L}{2}} \frac{\Delta}{\sigma}, 1\right)$, where Δ is the treatment difference and σ^2 is the population variance. Therefore the conditional power is

$$C_P = 1 - \Phi\left(c_K \sqrt{1 + \frac{n_L}{n_{\max} - n_L}} - Z_L \sqrt{\frac{n_L}{n_{\max} - n_L}} - \frac{\Delta \sqrt{m_{\max} - n_L}}{\sqrt{2}\sigma}\right).$$

Note that this formula is slightly different than the conditional power used in the EAST 5.1 manual [6]. In EAST 5.1, the denominator in the last portion is 2σ not $\sqrt{2}\sigma$. Below is the reason why it should be $\sqrt{2}\sigma$.

For simplicity, assume two normal populations $x_{1i} \sim N\left(\mu_1, \sigma^2\right)$, and $x_{2i} \sim N\left(\mu_2, \sigma^2\right)$. Let $\Delta = \mu_1 - \mu_2$, then $\left(\overline{x_1} - \overline{x_2}\right) \sim N\left(\Delta, \frac{2\sigma^2}{n}\right)$, and $S_n = \sum_{i=1}^{n} \left(x_{1i} - x_{2i}\right)$. Thus $Z_n = \frac{S_n}{\sqrt{2n\sigma^2}} \sim N\left(\sqrt{\frac{n}{2}} \frac{\Delta}{\sigma}, 1\right)$. So the denominator should be $\sqrt{2}\sigma$.

2.3.5 Calculation of Group Sequential Boundaries

The classical approach to calculate the boundaries is to solve the following equations by using the numeric method.

Under the null hypothesis, the Pocock boundaries satisfy

$$P_{H_0}\left(Z_{t_1} > c,\ or\ Z_{t_2} > c, \ldots\ or\ Z_{t_K} > c\right) = \alpha.$$

Under the null hypothesis and the O'Brien-Fleming boundaries satisfy

$$P_{H_0}\left(Z_{t_1}\sqrt{t_1} > c,\ or\ Z_{t_2}\sqrt{t_2} > c, \ldots\ or\ Z_{t_K}\sqrt{t_K} > c\right) = \alpha,$$

where $t_1, t_2, \ldots, t_K$ are the information fraction at each interim look. As discussed in section 2.3.2, boundaries can also be obtained approximately based on the alpha spending function approach. Because of the flexibility it provides, the alpha spending function approach is more often used in practice. For review purposes, the calculations based on the classical approach is presented in this section.

Density functions of test statistics at interim look k

Denote the (pseudo) density function $g_k\left(Z_k, \Delta\right)$ at kth interim look:

$$g_k\left(Z_k, \Delta\right) = f\left(Z_k = x, Z_1 < c_1, Z_2 < c_2, \ldots, Z_{k-1} < c_{k-1}\right).$$

At each look, this pseudo density function can be computed as:

1) At first interim look:

$g_1\left(Z_1, \Delta\right) = f\left(Z_1, \Delta\right)$, $f\left(Z_1, \Delta\right)$ is a normally distributed with mean $\sqrt{\frac{Nt_1}{2}}\Delta$ and a variance of 1. Thus $g_1\left(Z_1, \Delta\right) = \Phi\left(Z_1 - \sqrt{\frac{Nt_1}{2}}\Delta\right)$.

2) At second interim look:

Since

$$f\left(Z_2|Z_1,\Delta\right)\sim N\left(\sqrt{\frac{Nt_2}{2}}\Delta+\sqrt{\frac{t_1}{t_2}}\left(Z_1-\sqrt{\frac{Nt_1}{2}}\Delta\right),1-\frac{t_1}{t_2}\right),$$

$$\begin{aligned}P\left(Z_2<x,Z_1<c_1\right)&=\int_{-\infty}^{c_1}P\left(Z_2<x\right)g_1\left(Z_1,\Delta\right)dZ_1\\&=\int_{-\infty}^{c_1}P\left(\frac{\left(Z_2-\sqrt{\frac{Nt_2}{2}}\Delta\right)-\sqrt{\frac{t_1}{t_2}}\left(Z_1-\sqrt{\frac{Nt_1}{2}}\Delta\right)}{\sqrt{1-\frac{t_1}{t_2}}}\right.\\&\qquad\left.<\frac{\left(x-\sqrt{\frac{Nt_2}{2}}\Delta\right)-\sqrt{\frac{t_1}{t_2}}\left(Z_1-\sqrt{\frac{Nt_1}{2}}\Delta\right)}{\sqrt{1-\frac{t_1}{t_2}}}\right)g_1\left(Z_1,\Delta\right)dZ_1\\&=\int_{-\infty}^{c_1}\Phi\left(\frac{\left(x-\sqrt{\frac{Nt_2}{2}}\Delta\right)-\sqrt{\frac{t_1}{t_2}}\left(Z_1-\sqrt{\frac{Nt_1}{2}}\Delta\right)}{\sqrt{1-\frac{t_1}{t_2}}}\right)g_1\left(Z_1,\Delta\right)dZ_1.\end{aligned}$$

Taking derivative with respect to x, we get:

$$\begin{aligned}g_k\left(Z_k,\Delta\right)=&\int_{-\infty}^{c_{k-1}}\sqrt{\frac{t_k}{t_k-t_{k-1}}}\phi\left(\frac{\left(x-\sqrt{\frac{Nt_k}{2}}\Delta\right)-\sqrt{\frac{t_{k-1}}{t_k}}\left(Z_{k-1}-\sqrt{\frac{Nt_{k-1}}{2}}\Delta\right)}{\sqrt{1-\frac{t_{k-1}}{t_k}}}\right)\\&*\ g_{k-1}\left(Z_{k-1},\Delta\right)dZ_{k-1}.\end{aligned}$$

3) Similarly, at k^{th} interim look:

Since

$$f\left(Z_k|Z_{k-1},\Delta\right)\sim N\left(\sqrt{\frac{Nt_k}{2}}\Delta+\sqrt{\frac{t_{k-1}}{t_k}}\left(Z_{k-1}-\sqrt{\frac{Nt_{k-1}}{2}}\Delta\right),1-\frac{t_{k-1}}{t_k}\right),$$

$$\begin{aligned}&P\left(Z_k<x,Z_1<c_1,Z_2<c_2,...,Z_k<c_k\right)\\&=\int_{-\infty}^{c_{k-1}}P\left(Z_k<x\right)g_{k-1}\left(Z_{k-1},\Delta\right)dZ_{k-1}\end{aligned}$$

$$= \int_{-\infty}^{c_{k-1}} P\left(\frac{\left(Z_k - \sqrt{\frac{Nt_k}{2}}\Delta\right) - \sqrt{\frac{t_{k-1}}{t_k}}\left(Z_{k-1} - \sqrt{\frac{Nt_{k-1}}{2}}\Delta\right)}{\sqrt{1-\frac{t_{k-1}}{t_k}}} \right.$$

$$\left. < \frac{\left(x - \sqrt{\frac{Nt_k}{2}}\Delta\right) - \sqrt{\frac{t_{k-1}}{t_k}}\left(Z_{k-1} - \sqrt{\frac{Nt_{k-1}}{2}}\Delta\right)}{\sqrt{1-\frac{t_{k-1}}{t_k}}} \right) g_{k-1}\left(Z_{k-1}, \Delta\right) dZ_{k-1}$$

$$= \int_{-\infty}^{c_{k-1}} \Phi\left(\frac{\left(x - \sqrt{\frac{Nt_k}{2}}\Delta\right) - \sqrt{\frac{t_{k-1}}{t_k}}\left(Z_{k-1} - \sqrt{\frac{Nt_{k-1}}{2}}\Delta\right)}{\sqrt{1-\frac{t_{k-1}}{t_k}}} \right) g_{k-1}\left(Z_{k-1}, \Delta\right) dZ_{k-1}.$$

Taking derivative with respect to x, we get:

$$g_k\left(Z_k, \Delta\right) = \int_{-\infty}^{c_{k-1}} \sqrt{\frac{t_k}{t_k - t_{k-1}}} \phi\left(\frac{\left(x - \sqrt{\frac{Nt_k}{2}}\Delta\right) - \sqrt{\frac{t_{k-1}}{t_k}}\left(Z_{k-1} - \sqrt{\frac{Nt_{k-1}}{2}}\Delta\right)}{\sqrt{1-\frac{t_{k-1}}{t_k}}} \right)$$

$$* \; g_{k-1}\left(Z_{k-1}, \Delta\right) dZ_{k-1}.$$

Calculation of Boundaries: Under the null hypothesis, $\Delta = 0$. Two-sided symmetric Pocock boundaries can be calculated numerically by solving the following equation using the bi-section method. Integrations will be done by applying Simpson's rule

$$\alpha = P_{H_0}\left(Z_{t_1} > c, \; or \; Z_{t_2} > c, ... \; or \; Z_{t_K} > c\right)$$

$$= P_{H_0}\left(Z_1 > c\right) + P_{H_0}\left(Z_1 < c, Z_2 > c\right) + ... + P_{H_0}\left(Z_1 < c, Z_2 < c, ..., Z_{K-1} < c, Z_K > c\right)$$

$$= 1 - \Phi\left(c\right) + \int_{-\infty}^{c} \left(1 - \Phi\left(\frac{c - \sqrt{\frac{t_1}{t_2}} Z_1}{\sqrt{1-\frac{t_1}{t_2}}}\right)\right) g_1\left(Z_1, 0\right) dZ_1 + ...$$

$$+ \int_{-\infty}^{c} \left(1 - \Phi\left(\frac{c - \sqrt{\frac{t_{K-1}}{t_K}} Z_K}{\sqrt{1-\frac{t_{K-1}}{t_K}}}\right)\right) g_{K-1}\left(Z_{K-1}, 0\right) dZ_{K-1}.$$

Same as Pocock boundaries, O'Brien Fleming boundaries can be calculated by replacing the integration upper limit by $\frac{c}{\sqrt{t_i}}$, where t_i is the information fraction at each look.

2.3.6 Expected Sample Size for Group Sequential Designs

Under the alternative hypothesis $H_a : \Delta = \theta$, the expected sample size can be calculated. Boundaries are known now. The expected sample size is

$$E(N) = \sum_{k=1}^{K-1} P_{H_a}(Z_1 < c_1, Z_2 < c_2, ..., Z_{k-1} < c_k, Z_k > c_k) * n_k + n_K \left(1 - \sum_{k=1}^{K-1} P_{H_a}(Z_1 < c_1, Z_2 < c_2, ..., Z_{k-1} < c_k, Z_k > c_k)\right),$$

where N and n_k are the total sample size and the sample size at k^{th} look.

$P_{H_a}(Z_1 < c_1, Z_2 < c_2, ..., Z_{k-1} < c_k, Z_k > c_k)$ is the probability of stopping the trial at k^{th} look, we have

$$P_{H_a}(Z_1 < c_1, Z_2 < c_2, ..., Z_{k-1} < c_k, Z_k > c_k) = \int_{-\infty}^{c} 1 - \Phi\left(\frac{\left(c_k - \sqrt{\frac{Nt_k}{2}}\Delta\right) - \sqrt{\frac{t_{k-1}}{t_k}}\left(Z_{k-1} - \sqrt{\frac{Nt_{k-1}}{2}}\Delta\right)}{\sqrt{1 - \frac{t_{k-1}}{t_k}}}\right) g_{k-1}(Z_{k-1}, \Delta)\, dZ_{k-1}.$$

Chapter 3

Expected Sample Size for Two-stage Sample Size Re-estimation Designs

The expected sample size for group sequential designs can be calculated as specified in section 2.3.6. In this chapter, the expected sample size for weighted and unweighted sample size re-estimation designs with total 2 looks (two-stage sample size re-estimation designs) will be derived. For simplicity, it is assumed no futility test will be done at the interim analysis and no sample size reduction will be allowed.

3.1 Notation

The following notations will be used in this section:

n_1 and n_2 : Planned one arm sample size at first and second stage

$n_{2_{\max}}$: the maximum sample size allowed at stage 2 after sample size re-estimation

n_2^* : adjusted sample size at stage 2

t_1 : Information fraction at first stage, $t_1 = \dfrac{n_1}{n_1 + n_2}$

c_1 and c_2 : Boundary at first and second stage

Z_1 and Z_2 : Z-score at first and second stage

δ : True treatment effect

$\widehat{\delta}$: Estimated treatment effect at interim, $\widehat{\delta} = Z_1\sqrt{\dfrac{2}{n_1}}$

C_{P_0} : Targeted conditional power for sample size adjustment at interim look

$Z_{C_{P_0}}$: Corresponding Z-score for C_{P_0}

3.2 Weighted two-stage sample size re-estimation design

For the weighted sample size re-estimation design, the expected sample size is derived separately based on if there is a limitation for the maximum sample size.

3.2.1 With restriction of maximum sample size at second stage

For two-stage CHW design, the conditional power is computed as:

$$\begin{aligned} C_P &= P(Z_2 > c_2 | Z_1) \\ &= 1 - \Phi\left[\frac{c_2 - \sqrt{\frac{n_1+n_2^*}{2}}\widehat{\delta} - \sqrt{t_1}(Z_1 - \sqrt{\frac{n_1}{2}}\widehat{\delta})}{\sqrt{1-t_1}}\right] \\ &= 1 - \Phi\left[\frac{c_2 - \sqrt{\frac{n_1+n_2^*}{n_1}}Z_1}{\sqrt{1-t_1}}\right]. \end{aligned}$$

A more generalized conditional power formula for CHW design with multiple looks can be found in Appendix A.

When $C_P \geq C_{P_0}$, no sample size adjustment is needed ($n_2^* = n_2$)

$$C_P \geq C_{P_0}$$
$$\Leftrightarrow Z_1 \geq \sqrt{t_1\,(1-t_1)}Z_{C_{P_0}} + \sqrt{t_1}c_2.$$

Let $b = \sqrt{t_1\,(1-t_1)}Z_{C_{P_0}} + \sqrt{t_1}c_2$, then we have

$$C_P \geq C_{P_0} \Leftrightarrow Z_1 \geq b,$$

and

$$C_P < C_{P_0} \Leftrightarrow Z_1 < b.$$

When $C_P < C_{P_0}$, sample size will be adjusted. By setting $C_P = C_{P_0}$, we have

$$n_2^* = n_1\left[\left(\frac{\sqrt{1-t_1}Z_{C_{P_0}} + c_2}{Z_1}\right)^2 - 1\right].$$

When C_P is extremely small, say when $n_2^* > n_{2_{\max}}$, the adjusted sample size will be set to $n_{2_{\max}}$.

$$n_2^* = n_1\left[\left(\frac{\sqrt{1-t_1}Z_{C_{P_0}} + c_2}{Z_1}\right)^2 - 1\right] > n_{2_{\max}}$$
$$\Leftrightarrow Z_1 < \frac{\sqrt{1-t_1}Z_{C_{P_0}} + c_2}{\sqrt{\frac{n_1+n_{2\max}}{n_1}}}.$$

Let $a = \frac{\sqrt{1-t_1}Z_{C_{P_0}} + c_2}{\sqrt{\frac{n_1+n_{2\max}}{n_1}}}$, we have

$$n_2^* > n_{2_{\max}} \Leftrightarrow Z_1 < a.$$

Furthermore, we have

$$b = \sqrt{t_1\,(1-t_1)}Z_{C_{P_0}} + \sqrt{t_1}c_2$$

$$
\begin{aligned}
&= \sqrt{t_1\left(1-t_1\right)}\sqrt{\frac{n_1+n_{2\max}}{n_1}}Z_{C_{P_0}} + \sqrt{t_1}\sqrt{\frac{n_1+n_{2\max}}{n_1}}c_2 \\
&= \sqrt{\frac{n_1}{n_1+n_2}\frac{n_2}{n_1+n_2}}\sqrt{\frac{n_1+n_{2\max}}{n_1}}Z_{C_{P_0}} + \sqrt{\frac{n_1}{n_1+n_2}}\sqrt{\frac{n_1+n_{2\max}}{n_1}}c_2 \\
&= \sqrt{1-t_1}\sqrt{\frac{n_1+n_{2\max}}{n_1+n_2}}Z_{C_{P_0}} + \sqrt{\frac{n_1+n_{2\max}}{n_1+n_2}}c_2 \\
&= a\sqrt{\frac{n_1+n_{2\max}}{n_1+n_2}}.
\end{aligned}
$$

Since $\sqrt{\frac{n_1+n_{2\max}}{n_1+n_2}} \geq 1$, we have $b \geq a$. Hence the expected sample size can be computed as

$$
\begin{aligned}
E(N|\delta) &= 2n_1P(Z_1 > c_1|\delta) + 2\left(n_1+n_2\right)P(C_P \geq C_{P_0} \text{ and } Z_1 \leq c_1|\delta) \\
&\quad + \int\limits_{C_P<C_{P_0}\,\&\,z_1\leq c_1} 2\left(n_1+n_2^*\right)\phi\left(z_1|\delta\right)dz_1 \\
&= 2n_1P(Z_1 > c_1|\delta) + 2\left(n_1+n_2\right)P(b \leq Z_1 \leq c_1|\delta) \\
&\quad + \int\limits_{z_1\leq \min(b,c_1)} 2\left(n_1+n_2^*\right)\phi\left(z_1|\delta\right)dz_1 \\
&= 2n_1P(Z_1 > c_1|\delta) + 2\left(n_1+n_2\right)P\left(b \leq Z_1 \leq c_1|\delta\right) + 2(n_{1+}n_{2_{\max}})P\left(Z_1 < \min(a,c_1)|\delta\right) \\
&\quad + 2\int_a^b \left[n_1 + n_1\left[\left(\frac{\sqrt{1-t_1}Z_{C_{P_0}}+c_2}{z_1}\right)^2 - 1\right]\right]\phi\left(z_1|\delta\right)I(z_1 \leq c_1)dz_1 \\
&= 2n_1\left[P(Z_1 > c_1) + P\left(b \leq Z_1 \leq c_1\right) + P\left(Z_1 < \min(a,c_1)\right) + \int_a^b \phi\left(z_1|\delta\right)I(z_1 \leq c_1)dz_1\right] \\
&\quad + 2n_2P\left(b \leq Z_1 \leq c_1\right) + 2n_{2_{\max}}P\left(Z_1 < \min(a,c_1)\right) \\
&\quad + 2\int_a^b n_1\left[\left(\frac{\sqrt{1-t_1}Z_{C_{P_0}}+c_2}{Z_1}\right)^2 - 1\right]\phi\left(z_1|\delta\right)I(z_1 \leq c_1)dz_1 \\
&= 2n_1 + 2n_2P\left(b \leq Z_1 \leq c_1\right) + 2n_{2_{\max}}P\left(Z_1 < \min(a,c_1)\right) \\
&\quad + 2\int_a^b n_1\left[\left(\frac{\sqrt{1-t_1}Z_{C_{P_0}}+c_2}{Z_1}\right)^2 - 1\right]\phi\left(z_1|\delta\right)I(z_1 \leq c_1)dz_1
\end{aligned}
$$

Since the observed treatment difference follows a normal distribution with mean

δ and variance of 2, we have $Z_1 \sim N\left(\sqrt{\frac{n_1}{2}}\delta, 1\right)$.

Thus

$$
\begin{aligned}
E(N|\delta) &= 2n_1 P(Z_1 > c_1|\delta) + 2\left(n_1 + n_2\right) P(C_P \geq C_{P_0} \text{ and } Z_1 \leq c_1|\delta) \\
&\quad + \int\limits_{C_P < C_{P_0} \,\&\, z_1 \leq c_1} 2\left(n_1 + n_2^*\right) \phi\left(z_1|\delta\right) dz_1 \\
&= 2n_1 + 2n_2 \left[\Phi\left(c_1 - \sqrt{\frac{n_1}{2}}\delta\right) - \Phi\left(b - \sqrt{\frac{n_1}{2}}\delta\right)\right] I(b \leq c_1) \\
&\quad + 2n_{2_{\max}} \Phi\left(\min(a, c_1) - \sqrt{\frac{n_1}{2}}\delta\right) \\
&\quad + 2\int_a^b n_1 \left(\frac{\sqrt{1-t_1} Z_{C_{P_0}} + c_2}{z_1}\right)^2 \phi\left(z_1|\delta\right) I(z_1 \leq c_1) dz_1 \\
&\quad - 2\int_a^b n_1 \phi\left(z_1|\delta\right) I(z_1 \leq c_1) dz_1 \\
&= 2n_1 + 2n_2 \left[\Phi\left(c_1 - \sqrt{\frac{n_1}{2}}\delta\right) - \Phi\left(b - \sqrt{\frac{n_1}{2}}\delta\right)\right] I(b \leq c_1) \\
&\quad + 2n_{2_{\max}} \Phi\left(\min(a, c_1) - \sqrt{\frac{n_1}{2}}\delta\right) \\
&\quad - 2n_1 \left[\Phi\left(\min(b, c_1) - \sqrt{\frac{n_1}{2}}\delta\right) - \Phi\left(a - \sqrt{\frac{n_1}{2}}\delta\right)\right] I(a < c_1) \\
&\quad + 2n_1 \left(\sqrt{1-t_1} Z_{C_{P_0}} + c_2\right)^2 \int_a^b \left(\frac{1}{z_1}\right)^2 \phi\left(z_1|\delta\right) I(z_1 \leq c_1) dz_1, \qquad (3.1)
\end{aligned}
$$

Since there is no closed form formula for

$$
\begin{aligned}
&\int_a^b \left(\frac{1}{z_1}\right)^2 \phi\left(z_1|\delta\right) I(z_1 \leq c_1) dz_1 \\
&= \int_a^b \left(\frac{1}{z_1}\right)^2 \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{\left(z_1 - \sqrt{\frac{n_1}{2}}\delta\right)^2}{2}}\right) I(z_1 \leq c_1) dz_1,
\end{aligned}
$$

numeric method will be used to evaluate the integration. Simulations are done for two-stage sample size re-estimation designs and two-stage group sequential designs. Comparisons between theoretical calculation and simulation are presented in Tables 3.1 and 3.2 and Figure 3.1. Results from simulations are very close to the theoretical results.

Table 3.1: Comparison of expected sample size for sample size re-estimation designs with different boundaries*

	OBF		PK		HP01		HP005	
δ	simulation	calculation	simulation	calculation	simulation	calculation	simulation	calculation
0.0882	1389	1336	1411	1375	1386	1339	1374	1336
0.1294	1121	1083	1147	1116	1117	1084	1126	1083
0.1706	877	838	898	857	872	835	882	838
0.2117	649	629	675	629	654	621	655	629
0.2529	483	467	471	451	480	455	488	468
0.2941	360	353	334	328	349	340	363	354
0.3353	280	279	256	253	270	266	282	279
0.3765	235	232	215	212	224	222	235	233
0.4176	206	205	193	192	199	199	206	206
0.4588	191	191	184	183	187	187	191	191
0.5	183	183	180	180	181	181	183	183

* OBF: O'Brien-Fleming boundaries
* PK: Pocock boundaries
* HP01: Haybittle-Peto boundaries with $\alpha_0 = 0.01$
* HP005: Haybittle-Peto boundaries with $\alpha_0 = 0.005$

Table 3.2: Comparison of expected sample size for group sequential designs with different boundaries*

	OBF		PK		HP01		HP005	
δ	simulation	calculation	simulation	calculation	simulation	calculation	simulation	calculation
0.0882	1751	1809	1461	1592	1668	1740	1754	1812
0.1294	1461	1469	1231	1244	1375	1383	1467	1474
0.1706	1154	1161	1057	1059	1111	1114	1156	1163
0.2117	1034	1034	1014	1014	1023	1024	1035	1035
0.2529	1012	1011	1010	1009	1011	1010	1012	1011
0.2941	1009	1009	1009	1009	1009	1009	1009	1009
0.3353	1009	1009	1009	1009	1009	1009	1009	1009
0.3765	1009	1009	1009	1009	1009	1009	1009	1009
0.4176	1009	1009	1009	1009	1009	1009	1009	1009
0.4588	1009	1009	1009	1009	1009	1009	1009	1009
0.5	1009	1009	1009	1009	1009	1009	1009	1009

* OBF: O'Brien-Fleming boundaries
* PK: Pocock boundaries
* HP01: Haybittle-Peto boundaries with $\alpha_0 = 0.01$
* HP005: Haybittle-Peto boundaries with $\alpha_0 = 0.005$

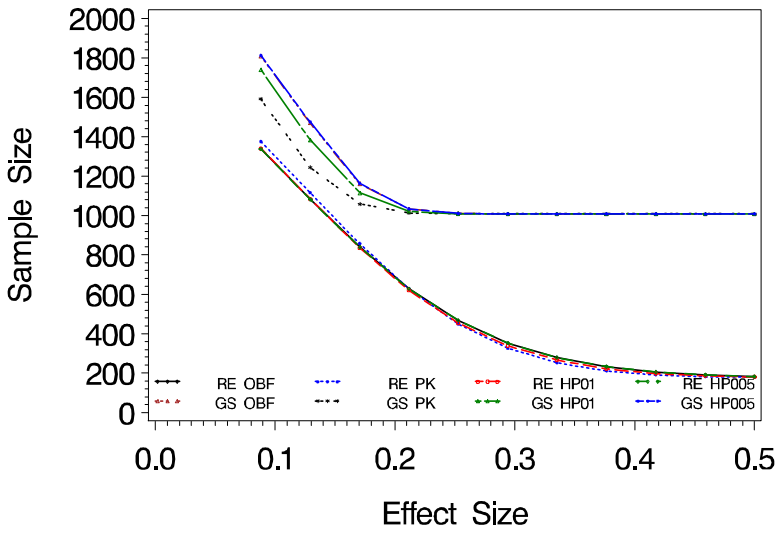

Figure 3.1: Expected Sample Size for Two-stage GS Designs and Weighted Sample Size Re-estimation Designs

3.2.2 Without restriction of maximum sample size at the second stage

Method 1: Since

$$\lim_{n_{2\max}\to\infty} a = 0,$$

where

$$a = \frac{\sqrt{1-t_1}Z_{C_{P_0}} + c_2}{\sqrt{\frac{n_1+n_{2\max}}{n_1}}},$$

and $b = \left(\sqrt{1-t_1}Z_{C_{P_0}} + c_2\right)\sqrt{t_1}$ does not depend on $n_{2_{\max}}$, all parts in equation 3.1 are finite except $2n_{2_{\max}}\Phi\left(\min(a, c_1) - \sqrt{\frac{n_1}{2}}\delta\right)$ which will go to infinity when $n_{2_{\max}}$ goes to infinity. Thus, when $n_{2\max} \to \infty$, the expected sample size will go to infinity.

Method 2:

$$\begin{aligned}
E(N|\delta) &= 2n_1P(Z_1 > c_1|\delta) + 2(n_1+n_2)P(C_P \geq C_{P_0}\text{and } Z_1 \leq c_1|\delta) \\
&\quad + \int\limits_{C_P<C_{P_0}\,\&\, Z_1\leq c_1} 2(n_1+n_2^*)\phi(z_1|\delta)\,dz_1 \\
&= 2n_1P(Z_1 > c_1|\delta) + 2(n_1+n_2)P(b \leq Z_1 \leq c_1|\delta) + \int\limits_{z_1\leq \min(b,c_1)} 2(n_1+n_2^*)\phi(z_1|\delta)\,dz_1 \\
&= 2n_1P(Z_1 > c_1|\delta) + 2(n_1+n_2)P(b \leq Z_1 \leq c_1|\delta) \\
&\quad + 2\int_{-\infty}^{b}\left[n_1 + n_1\left[\left(\frac{\sqrt{1-t_1}Z_{C_{P_0}} + c_2}{z_1}\right)^2 - 1\right]\right]\phi(z_1|\delta)\,I(z_1 \leq c_1)dz_1 \\
&= 2n_1\left[P(Z_1 > c_1|\delta) + P(b \leq Z_1 \leq c_1|\delta) + \int_{-\infty}^{b}\phi(z_1|\delta)\,I(z_1 \leq c_1)dz_1\right] \\
&\quad + 2n_2P(b \leq Z_1 \leq c_1|\delta) + 2\int_{-\infty}^{b} n_1\left[\left(\frac{\sqrt{1-t_1}Z_{C_{P_0}} + c_2}{z_1}\right)^2 - 1\right]\phi(z_1|\delta)\,I(z_1 \leq c_1)dz_1 \\
&= 2n_1 + 2n_2\left[\Phi\left(c_1 - \sqrt{\frac{n_1}{2}}\delta\right) - \Phi\left(b - \sqrt{\frac{n_1}{2}}\delta\right)\right]I(b \leq c_1) \\
&\quad + 2n_1\left(\sqrt{1-t_1}Z_{C_{P_0}} + c_2\right)^2\int_{-\infty}^{b}\left(\frac{1}{z_1}\right)^2\phi(z_1|\delta)\,I(z_1 \leq c_1)\,dz_1
\end{aligned}$$

$$
\begin{aligned}
&- 2n_1 \int_{-\infty}^{b} \phi\left(z_1|\delta\right) I\left(z_1 \leq c_1\right) dz_1 \\
&= 2n_1 + 2n_2 \left[\Phi\left(c_1 - \sqrt{\frac{n_1}{2}}\delta\right) - \Phi\left(b - \sqrt{\frac{n_1}{2}}\delta\right)\right] I(b \leq c_1) \\
&\quad - 2n_1 \Phi\left(\min(b, c_1) - \sqrt{\frac{n_1}{2}}\delta\right) \\
&\quad + 2n_1 \left(\sqrt{1-t_1} Z_{C_{P_0}} + c_2\right)^2 \int_{-\infty}^{b} \left(\frac{1}{z_1}\right)^2 \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{\left(z_1 - \sqrt{\frac{n_1}{2}}\delta\right)^2}{2}}\right) I\left(z_1 \leq c_1\right) dz_1,
\end{aligned}
$$

where

$$
b = \sqrt{t_1\left(1-t_1\right)} Z_{C_{P_0}} + \sqrt{t_1} c_2.
$$

Since $\frac{1}{Z_1^2}$ diverges for integral include 0,

$$
\int_{-\infty}^{b} \left(\frac{1}{z_1}\right)^2 \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{\left(z_1 - \sqrt{\frac{n_1}{2}}\delta\right)^2}{2}}\right) I\left(z_1 \leq c_1\right) dz_1 = \infty.
$$

When $n_{2\max} \to \infty$, the expected sample size will go to infinity. So method 1 and method 2 led to the same conclusion that the expected sample size will go to infinity when there is no restriction on the sample size allowed in the second stage.

3.3 Unweighted two-stage sample size re-estimation design

In weighted two-stage design, the adjusted sample size n_2^* for stage 2 is calculated by solving the equation

$$
C_P = C_{P_0},
$$

where

$$
C_P = 1 - \Phi\left[\frac{c_2 - \sqrt{\frac{n_1 + n_2^*}{n_1}} Z_1}{\sqrt{1-t_1}}\right],
$$

and

$$n_2^* = n_1 \left[\left(\frac{\sqrt{1-t_1} Z_{C_{P_0}} + c_2}{Z_1} \right)^2 - 1 \right],$$

where c_2 is the boundary at stage 2 determined by the initial sample size and the information fraction at stage 1.

The calculation of expected sample size for unweighted design will depend on how the conditional power (C_P) is calculated. One way to calculate C_P is the same as in the weighted design,

$$C_P = 1 - \Phi \left[\frac{c_2 - \sqrt{\frac{n_1+n_2^*}{n_1}} Z_1}{\sqrt{1-t_1}} \right],$$

where c_2 is the boundary at stage 2 determined by the initial sample size and the information fraction at the interim analysis. In this case, because the adjusted sample size is not calculated based on the adjusted boundary at stage 2, the calculation of expected sample size will be the same as the weighted method.

Another way to calculate C_P is to use the adjusted boundary c_2^* and

$$C_P = 1 - \Phi \left[\frac{c_2^* - \sqrt{\frac{n_1+n_2^*}{n_1}} Z_1}{\sqrt{1-t_1}} \right].$$

However, in order to calculate the expected sample size, certain relationship between c_2^* and n_2^* has to be assumed.

Chapter 4

Methods of Evaluation of Adaptive Design Performance

4.1 Determination of treatment effect interval

An appropriately selected treatment effect interval will have the robustness against the unknown true treatment effect in terms of average sample size and power. Treatment effect interval should be determined based on the combination of multiple considerations. Determination of the lower limit of treatment effect interval should be based on (1) clinical meaningful treatment difference, (2) medical policies, such as restriction of medication price, and (3) company financial considerations. Determination of upper limit of treatment effect interval should be based on (1) the minimum number of patients needed for adequate safety evaluation of the test drug and (2) a realistic estimate of the largest treatment difference.

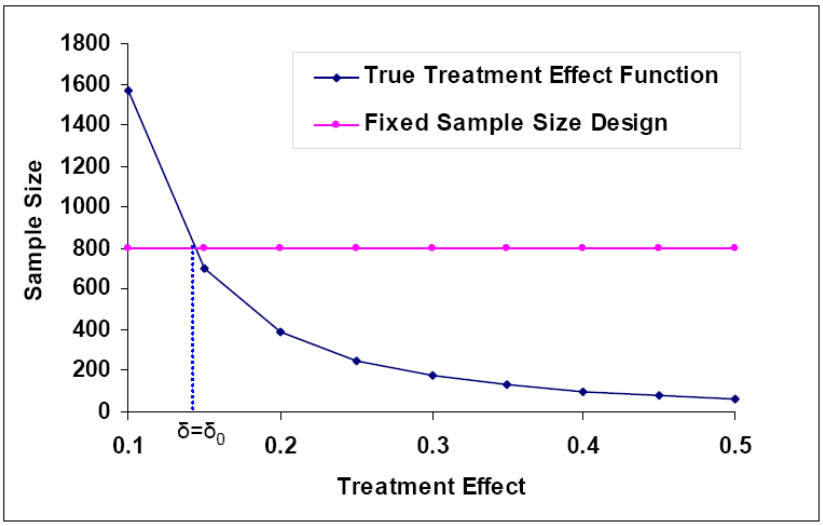

Figure 4.1: True Treatment Effect Function and Fixed Sample Size Design

4.2 Golden Standard for Performance Comparisons

In order to compare the performance for different adaptive designs, a function of the true treatment effect δ on the treatment effect interval $[\delta_L, \delta_U]$ is used as the golden standard. When comparing the sample size, this function is

$$u(\delta) = \frac{2\left(z_\beta + z_{\alpha/2}\right)^2}{\delta^2},$$

where α and β are the pre-specified type I error and type II error for the study. Sample size curves for the true treatment effect and for the fixed sample size design are illustrated in Figure 4.1. When comparing the power, this function is

$$p(\delta) = 1 - \beta.$$

4.3 Relationship between the True Treatment Effect Function and the Fixed Sample Size Design

Without knowing the true treatment effect δ, the sample size for fixed sample size design has to be calculated based on an estimated effect $\widehat{\delta}$ from a previous study or historical data. Sample size cannot be changed during the study. Only when $\widehat{\delta}$ is equal to true treatment effect δ, the fixed sample size design will have the ideal performance. When $\widehat{\delta}$ is shifted away from the δ, the fixed sample size design will be either under powered (if $\widehat{\delta} < \delta$) or oversized (if $\widehat{\delta} > \delta$). Its performance will be much worse than the performance of true treatment effect function at all other non-true treatment effects.

4.4 Measurements of Performance

4.4.1 Regret

Regret function in decision theory can be used to measure the performance of a design at the true treatment effect δ on the interval. Define regret as

$$l\left(\widehat{S}\left(\delta\right);S\left(\delta\right)\right)=\left|\widehat{S}\left(\delta\right)-S\left(\delta\right)\right|,\delta\in\left[\delta_{L},\delta_{U}\right],$$

where $S\left(\delta\right)$ is the sample size or power at the true δ and $\widehat{S}\left(\delta\right)$ is the estimated sample size or the power based on the design.

Regret for sample size is defined as the difference between the average sample size of adaptive design and the sample size from the true treatment effect function at δ. Regret for power is defined as the difference between the average power of adaptive design

and the power from the true treatment effect function at δ. The smaller the Regret is, the better the adaptive design performance is.

4.4.2 Failure Rate

Regret only measures the performance of adaptive design at a particular point on a treatment effect interval. It is important to identify a criterion to measure the cumulative performance on the treatment effect interval.

Failure and Failure Rate

At each point on the treatment effect interval, it is considered a failure when the sample size for a particular design at that point is more than $\frac{1}{f_s}$ (usually $\frac{1}{f_s} = 2$) times the sample size based on the true treatment effect or the power decreased more than f_p (usually 20%) of the power for the true treatment effect, where $0 < f_s < 1$ and $0 < f_p < 1$. The failure rate is defined as the proportion of points that meet the failure criteria on the treatment effect interval. A lower failure rate indicates a better performance of the adaptive design.

Failure Rate for Fixed Sample Size Design

Denote the sample size at the true treatment effect δ as $U(\delta)$ and n_0 the sample size calculated based on the fixed sample size design. Based on the above criteria, a failure occurs when n_0 is larger than $\frac{1}{f_s}U(\delta)$. Because the sample size curve for the true treatment effect is monotone, it is very easy to see that failure occurs for all $\delta > \delta_{fn}$ on the interval, where $\delta_{fn} = \sqrt{\frac{2\left(Z_{\alpha/2}+Z_{\beta}\right)^2}{f_s n_0}}$.

The targeted power on the interval for true treatment effect function is always $1-\beta$. Using the above criteria, it is a failure when power is decreased more than f_p times of the targeted power. Also, because of the monotonicity of power curves given the sample size n_0 calculated based on the fixed sample size design, it can be seen that the failure occurs for all $\delta < \delta_{f\beta}$ on the interval, where $\delta_{f\beta} = \sqrt{\frac{2\left(Z_{\alpha/2}+Z_{1-fp(1-\beta)}\right)^2}{n_0}}$.

Thus the total failure rate on a treatment effect interval is defined as:

$$R_f = \frac{\left(\delta_U - \delta_{fn}\right) + \left(\delta_{f\beta} - \delta_L\right)}{\delta_U - \delta_L} \times 100\%.$$

Below is an example of failure rate for fixed sample designs. When treatment effect interval is $[0.0882, 0.5]$ and assume $f_s = \frac{1}{2}$ and $f_p = 0.2$, the failure rates at different sample sizes for fixed designs are

1) Failure rate for fixed sample size design with sample size calculation at δ_U:

$R_f = (0.4131 - 0.0882)/(0.5 - 0.0882) \times 100\% = 78.9\%$. As indicated in Figure 4.2, when treatment effect falls into the region (0.0882$< \delta <$0.4131), power decreases more than 20% of the true treatment effect function for power $(1-\beta = 80\%)$.

2) Failure rate for fixed sample size design with sample size calculation at δ_L:

$R_f = (0.5 - 0.1247)/(0.5 - 0.0882) \times 100\% = 91.1\%$. As indicated in Figure 4.3, when treatment effect falls into the region ($0.1247 < \delta <$0.5), sample size is more than two times of the sample size from the true treatment function.

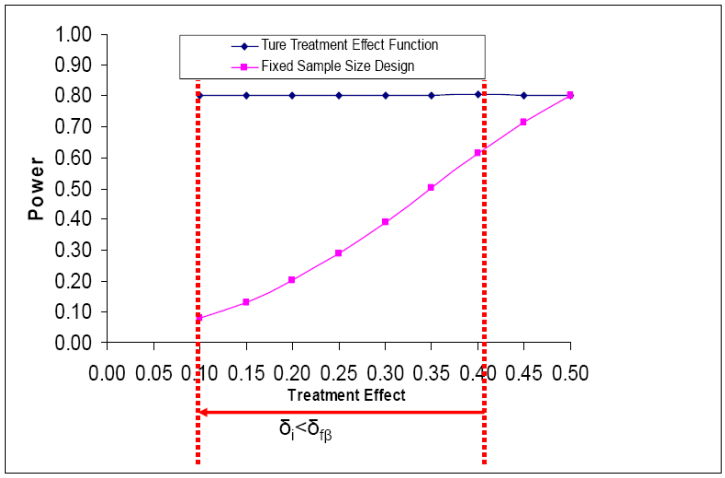

Figure 4.2: failure Rate of Power for Fixed Sample Size Design at δ_U

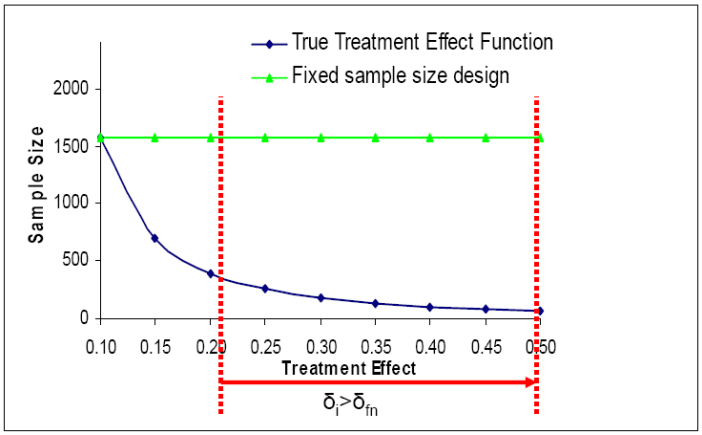

Figure 4.3: Failure Rate of Sample Size for Fixed Sample Size Design at δ_L

3) Failure rate for fixed sample size design with sample size calculation at $\delta_M = \sqrt{\delta_L \delta_U}$:

$R_f = [(0.5 - 0.2968) + (0.1738 - 0.0882)]/(0.5 - 0.0882) * 100\% = 70.1\%$. As indicated in Figure 4.4 and Figure 4.5, when treatment difference falls into the red region ($0.0882 < \delta < 0.1738$), power decreases more than 20% of the power from the true treatment effect function; when treatment difference falls into the red region ($0.2968 < \delta < 0.5$), sample size is more than two times of the sample size from the true treatment effect function.

Generalization of Failure Rate

A generalized formula for failure rate on a treatment effect interval is:

$$R_f = \int_{\delta_L}^{\delta_U} 1_{\left(\frac{g(x)}{u(x)} > \frac{1}{fs} \text{ or } 1 - \frac{f(x)}{p(x)} > f_p\right)}(x)\, W(x)\, dx,$$

where $g(x)$ and $f(x)$ denote the sample size and power for adaptive design when treatment effect is x, and $u(x)$ and $p(x)$ denote the sample size and power from the true treatment effect function when treatment effect is x. $W(x)$ denotes the weight assigned to $\delta \in [\delta_L, \delta_U]$ which will be a probability density function on the treatment effect interval. When treatment effects have a uniform distribution, $W(x)$ will be equal to $1/(\delta_U - \delta_L)$. Treatment effect also can be assumed to follow other distributions. In the later sections, treatment effect will be considered to follow a uniform distribution or a beta distribution.

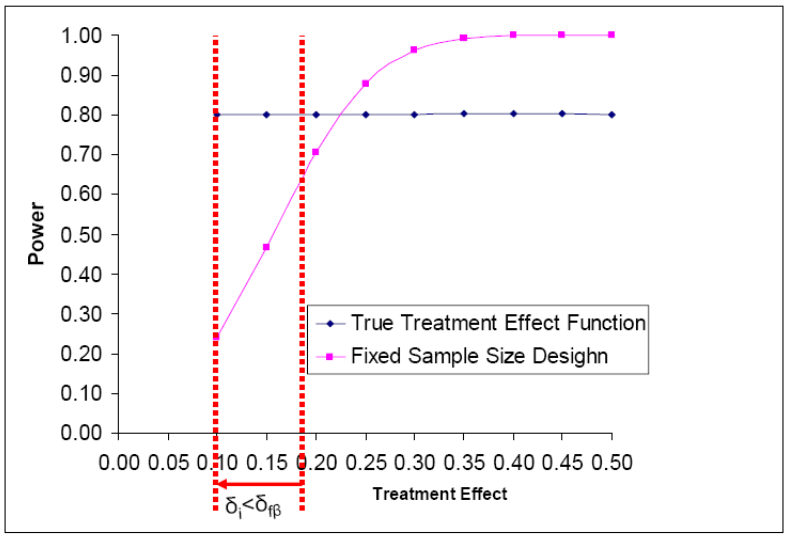

Figure 4.4: Failure Rate of Power for Fixed Sample Size Design at δ_M

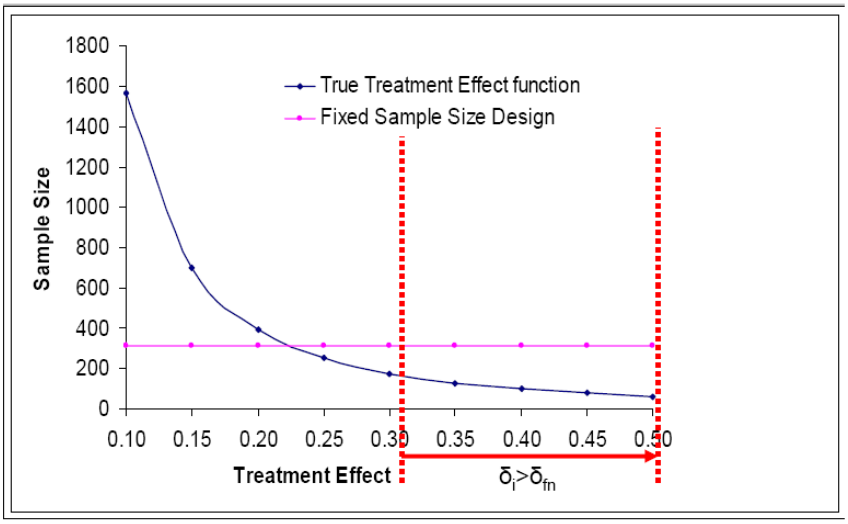

Figure 4.5: Failure Rate of Sample Size for Fixed Sample Size Design at δ_M

4.4.3 Area Between Curves (ABC) of Adaptive Design and True Treatment Effect Function

To measure the deviation from the true treatment effect function, ABC for each adaptive design and true treatment effect function can be calculated on the treatment effect interval (see Figure 4.6). Define ABC as

$$ABC = \int_{\delta_L}^{\delta_U} |g(x) - u(x)| \, dx.$$

Performance can be evaluated by comparing ABC for different designs. The smaller the area between the curves is, the better the performance is.

Area between Log Curves (ABLC) for Adaptive Design and True Treatment Effect Function

One can interpret $g(x) - u(x)$ as the absolute sample size difference between the adaptive design and the true treatment effect function. However, since the ratio of $g(x)$ to $u(x)$ measures the relative difference, it is more appropriate to be used to measure the deviation from the golden standard. For easier interpretation, log ratio can be calculated. We have

$$\log \frac{g(x)}{u(x)} = \log g(x) - \log u(x)$$

Define ABLC as

$$ABLC = \int_{\delta_L}^{\delta_U} | \log(g(x)) - \log(u(x))| \, dx.$$

ABLC will be used to compare the performance in the next sections.

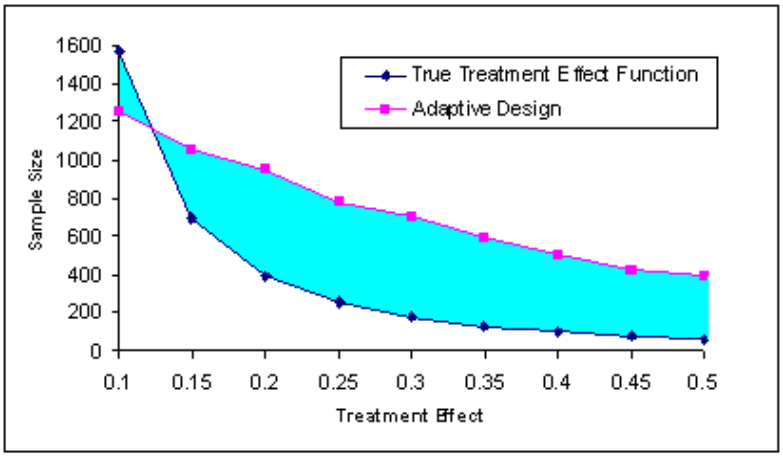

Figure 4.6: Area Between Curves

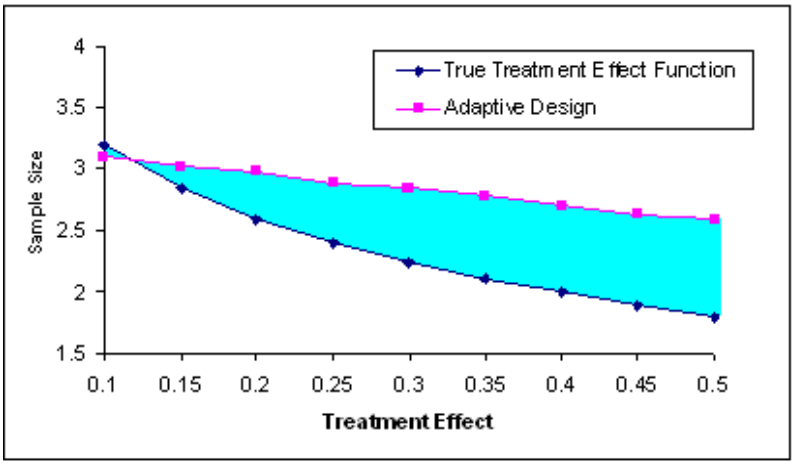

Figure 4.7: Area Between Log Curves

Chapter 5

Performance of Adaptive Designs when Treatment Effect Follows a Uniform Distribution

5.1 Method

In this chapter, the performance of group sequential designs, weighted sample size re-estimation designs (CHW method), and unweighted sample size re-estimation designs on an treatment effect interval are evaluated. It is assumed that treatment effect follows a uniform distribution. Thus, the probability of observing a treatment effect is the same on the treatment effect interval.

For group sequential designs, under a pre-specified total number of looks and maximum sample size, study is allowed to stop for efficacy at the interim analysis. CHW

design is used as a representation of weighted design. For design selected for unweighted method, boundaries for the interim looks after the sample size re-estimation are adjusted to maintain the overall type I error rate based on the original information time. For weighted and unweighted sample size re-estimation designs, an initial sample size is selected based on the sample size at δ_L, δ_U,and δ_M. Sample size can be increased based on the interim finding on treatment effect size during the study. But, there is a restriction for the maximum allowed sample size. Sample size re-estimation can be done based on the conditional power at selected look before the final analysis. Effects on the patient increment types and different adaptive index are studied. Performance scores under different patient increment or adaptive index are compared.

Four types of boundaries - O'Brien and Fleming type boundaries (OBF), Pocock type boundaries (PK), Haybittle-Peto boundaries with critical value of 0.01 (HP01), and Haybittle-Peto boundaries with critical value of 0.005 (HP005) are used in the performance comparisons (see table below). For group sequential designs and weighted sample size re-estimation designs, boundaries at each interim look and final look are fixed. However, for unweighted sample size re-estimation design, since sample size will be updated based on the interim finding, the information time needs to be updated. Thus, the boundaries after the sample size re-estimation look needs to be recalculated as well.

The effect of different patient increment patterns are also studied. Performance is compared when patient increment is equal-spaced or unequal-spaced with 2 time increments. Information time for each analysis is calculated based on the total number of looks and the types of increment as shown in Tables 5.1 and 5.2.

Table 5.1: Boundaries Used for Equal-spaced Designs

Total Looks	Method	Look 1	Look 2	Look 3	Look 4	Look 5	Look 6
2	*Information Time*	*0.5000*	*1.0000*				
	OBF	2.7965	1.9774				
	Pocock	2.1783	2.1783				
	HP01	2.5758	2.0027				
	HP005	2.8070	1.9767				
3	*Information Time*	*0.3333*	*0.6667*	*1.0000*			
	OBF	3.4711	2.4544	2.0040			
	Pocock	2.2895	2.2895	2.2895			
	HP01	2.5758	2.5758	2.0458			
	HP005	2.8070	2.8070	1.9933			
4	*Information Time*	*0.2500*	*0.5000*	*0.7500*	*1.0000*		
	OBF	4.0486	2.8628	2.3375	2.0243		
	Pocock	2.3613	2.3613	2.3613	2.3613		
	HP01	2.5758	2.5758	2.5758	2.0897		
	HP005	2.8070	2.8070	2.8070	2.0096		
5	*Information Time*	*0.2000*	*0.4000*	*0.6000*	*0.8000*	*1.0000*	
	OBF	4.5617	3.2256	2.6337	2.2809	2.0401	
	Pocock	2.4132	2.4132	2.4132	2.4132	2.4132	
	HP01	2.5758	2.5758	2.5758	2.5758	2.1349	
	HP005	2.8070	2.8070	2.8070	2.8070	2.0251	
6	*Information Time*	*0.1667*	*0.3333*	*0.5000*	*0.6667*	*0.8333*	*1.0000*
	OBF	5.0283	3.5555	2.9031	2.5142	2.2487	2.0528
	Pocock	2.4532	2.4532	2.4532	2.4532	2.4532	2.4532
	HP01	2.5758	2.5758	2.5758	2.5758	2.5758	2.1823
	HP005	2.8070	2.8070	2.8070	2.8070	2.8070	2.0400

Table 5.2: Boundaries Used for Unequal-spaced Designs

Total Looks	Method	Look 1	Look 2	Look 3	Look 4	Look 5	Look 6
3	*Information Time*	*0.2500*	*0.5000*	*1.0000*			
	OBF	3.9552	2.7967	1.9776			
	Pocock	2.3118	2.3118	2.3118			
	HP01	2.5758	2.5758	2.0714			
	HP005	2.8070	2.8070	2.0058			
4	*Information Time*	*0.1250*	*0.2500*	*0.5000*	*1.0000*		
	OBF	5.5935	3.9552	2.7967	1.9776		
	Pocock	2.4085	2.4085	2.4085	2.4085		
	HP01	2.5758	2.5758	2.5758	2.1599		
	HP005	2.8070	2.8070	2.8070	2.0408		
5	*Information Time*	*0.0625*	*0.1250*	*0.2500*	*0.5000*	*1.0000*	
	OBF	7.9104	5.5935	3.9552	2.7967	1.9776	
	Pocock	2.4843	2.4843	2.4843	2.4843	2.4843	
	HP01	2.5758	2.5758	2.5758	2.5758	2.2751	
	HP005	2.8070	2.8070	2.8070	2.8070	2.0803	
6	*Information Time*	*0.0313*	*0.0625*	*0.1250*	*0.2500*	*0.5000*	*1.0000*
	OBF	11.1870	7.9104	5.5935	3.9552	2.7967	1.9776
	Pocock	2.5464	2.5464	2.5464	2.5464	2.5464	2.5464
	HP01	2.5758	2.5758	2.5758	2.5758	2.5758	2.4407
	HP005	2.8070	2.8070	2.8070	2.8070	2.8070	2.1244

5.2 Simulation Plan

Simulations for adaptive designs will be based on the steps outlined below:

Step 1: Identify an interval of exploration and the maximum and minimum allowed sample size on the basis of early study results and literature.

Step 2: Choose candidate adaptive designs to be considered - group sequential designs with O'Brien Fleming boundary, Pocock boundary, Haybittle-Peto type boundary or sample size re-estimation designs using CHW method etc.

Step 3: Determine the following design parameters:

1. Adaptive index
2. Maximum sample size for group sequential designs and initial sample size for sample size re-estimation designs
3. Total number of looks
4. Types of information increment
5. Time of sample size re-estimation for sample size re-estimation designs
6. Adjustment of sample size at the predetermined interim look

Step 4: Obtain average sample size and power for 11 evenly spread treatment effects on the selected treatment effect interval for each design via Monte Carlo simulations

Step 5: Get the sample size and power curve on the treatment effect interval through interpolation

Step 6: Measure the performance for each adaptive design

All simulation results are based on 10000 runs for each treatment effect. Simulations are repeated based on different simulation parameters specified in step 3.

5.3 Results

5.3.1 Performance of adaptive designs

Performance for each adaptive design on the treatment effect interval $[0.0882, 0.5]$ was measured by failure rate and log area between curves. For group sequential design, performance is evaluated at different maximum allowed sample sizes - 2018, 356, and 63 which are the sample sizes in fixed sample size design when treatment effects are δ_L, δ_M, and δ_U. For sample size re-estimation designs, 356 is used as initial sample size (n_{init}) and the maximum allowed sample size after sample size re-estimation is 2018. Sample size re-estimation is based on a targeted conditional power of 80%. The method for calculating the new sample size is specified in section 3.2.

Performance scores for equal-spaced (patient increment) group sequential designs, weighted sample size re-estimation designs, and unweighted sample size re-estimation designs are presented in Tables 5.3, 5.5, and 5.7 respectively. Performance for designs with unequal-spaced 2-times increment are displayed in Tables 5.4, 5.6, and 5.8. Sample size and power curves for selected adaptive designs when treatment effect follows a uniform distribution are displayed in Appendix B.

In general for group sequential designs, when the maximum allowed sample size ($n_{\max}$) is 356, the performance scores are better than the scores when n_{max} is 63 or 2018.

There is a exception when the unequal-spaced patient increment is used. The best performance is observed when n_{max} is 2018. For weighted method, the best performance is obtained when the initial sample size is 356. When the initial sample size is 63, the sample size re-estimation is done based on very limited information and is not reliable. Thus, the performance scores are low. When the initial sample size is 2018, since no sample size re-estimation could be done, the sample size re-estimation designs become group sequential designs. Based on these observations on the weighted designs, unweighted designs are done only when initial sample size is 356.

Table 5.3: Performance Score Results for Equal Spaced GS Designs

		n_{max}=63		n_{max}=356		n_{max}=2018	
Total Looks	Method	R_f	ABLC	R_f	ABLC	R_f	ABLC
2	OBF	0.80	0.56	**0.50**	**0.27**	0.82	0.65
	Pocock	0.86	0.64	**0.45**	**0.27**	0.79	0.64
	HP01	0.81	0.58	**0.46**	**0.27**	0.81	0.65
	HP005	0.80	0.56	**0.50**	**0.27**	0.82	0.65
3	OBF	0.81	0.58	**0.46**	**0.26**	0.81	0.54
	Pocock	0.88	0.69	**0.25**	**0.24**	0.70	0.51
	HP01	0.81	0.61	**0.25**	**0.23**	0.72	0.52
	HP005	0.81	0.58	**0.29**	**0.23**	0.74	0.52
4	OBF	0.81	0.60	**0.42**	**0.26**	0.79	0.49
	Pocock	0.89	0.72	**0.25**	**0.22**	0.62	0.42
	HP01	0.83	0.63	**0.22**	**0.21**	0.64	0.43
	HP005	0.80	0.59	**0.21**	**0.22**	0.66	0.44
5	OBF	0.80	0.61	**0.37**	**0.25**	0.76	0.46
	Pocock	0.91	0.73	**0.26**	**0.22**	0.56	0.36
	HP01	0.83	0.65	**0.23**	**0.20**	0.57	0.37
	HP005	0.80	0.60	**0.22**	**0.21**	0.61	0.38
6	OBF	0.80	0.62	**0.33**	**0.25**	0.75	0.46
	Pocock	0.92	0.75	**0.26**	**0.22**	0.50	0.32
	HP01	0.85	0.67	**0.24**	**0.20**	0.51	0.32
	HP005	0.82	0.61	**0.22**	**0.20**	0.54	0.33

Table 5.4: Performance Score Results for Unequal Spaced 2-times Increment GS Designs

Total Looks	Method	n_{max}=63		n_{max}=356		n_{max}=2018	
		R_f	ABLC	R_f	ABLC	R_f	ABLC
3	OBF	0.80	0.56	**0.46**	**0.26**	0.81	0.49
	Pocock	0.90	0.69	**0.26**	**0.22**	0.62	0.42
	HP01	0.83	0.61	**0.22**	**0.21**	0.64	0.43
	HP005	0.80	0.57	**0.21**	**0.23**	0.67	0.44
4	OBF	0.79	0.56	**0.46**	**0.26**	0.81	0.47
	Pocock	0.93	0.72	**0.27**	**0.20**	0.39	0.25
	HP01	0.85	0.64	**0.24**	**0.20**	0.40	0.27
	HP005	0.82	0.59	**0.22**	**0.21**	0.44	0.29
5	OBF	0.79	0.56	**0.46**	**0.26**	0.81	0.47
	Pocock	0.96	0.74	**0.28**	**0.21**	0.05	0.15
	HP01	0.90	0.67	**0.25**	**0.20**	0.07	0.17
	HP005	0.83	0.60	**0.23**	**0.21**	0.14	0.21
6	OBF	0.80	0.56	**0.46**	**0.26**	0.81	0.47
	Pocock	0.99	0.75	**0.29**	**0.21**	0.00	0.12
	HP01	0.95	0.71	**0.28**	**0.21**	0.00	0.13
	HP005	0.85	0.61	**0.23**	**0.21**	0.00	0.18

Table 5.5: Performance Score Results for Equal Spaced Weighted Sample Size Re-estimation Designs

Total Look	Reestimation Look	Method	n_init=63		n_init=356		n_init=2018	
			R_f	LABC	R_f	LABC	R_f	LABC
2	1	OBF	0.90	0.44	**0.44**	**0.26**	0.82	0.65
		Pocock	0.93	0.45	**0.31**	**0.25**	0.79	0.64
		HP01	0.90	0.44	**0.36**	**0.26**	0.81	0.65
		HP005	0.89	0.43	**0.51**	**0.27**	0.82	0.65
3	1	OBF	0.67	0.29	**0.39**	**0.23**	0.81	0.54
		Pocock	0.71	0.30	**0.07**	**0.17**	0.70	0.51
		HP01	0.68	0.30	**0.10**	**0.19**	0.71	0.51
		HP005	0.71	0.31	**0.17**	**0.21**	0.74	0.52
	2	OBF	0.94	0.41	**0.32**	**0.23**	0.81	0.54
		Pocock	0.99	0.44	**0.07**	**0.18**	0.70	0.51
		HP01	0.94	0.41	**0.09**	**0.19**	0.72	0.52
		HP005	0.94	0.41	**0.13**	**0.20**	0.74	0.52
4	1	OBF	0.54	0.24	**0.32**	**0.21**	0.79	0.49
		Pocock	0.54	0.23	**0.06**	**0.14**	0.63	0.42
		HP01	0.50	0.23	**0.05**	**0.15**	0.64	0.43
		HP005	0.53	0.24	**0.04**	**0.18**	0.67	0.44
	2	OBF	0.68	0.27	**0.27**	**0.20**	0.79	0.49
		Pocock	0.83	0.30	**0.06**	**0.13**	0.63	0.42

Table 5.5: Performance Score Results for Equal Spaced Weighted Sample Size Re-estimation Designs

		HP01	0.84	0.29	**0.05**	**0.14**	0.64	0.43
		HP005	0.85	0.30	**0.04**	**0.17**	0.67	0.44
	3	OBF	0.96	0.40	**0.27**	**0.21**	0.79	0.49
		Pocock	1.03	0.45	**0.08**	**0.15**	0.63	0.42
		HP01	0.98	0.41	**0.06**	**0.15**	0.64	0.43
		HP005	0.96	0.40	**0.06**	**0.17**	0.66	0.44
5	1	OBF	0.55	0.23	**0.28**	**0.21**	0.76	0.46
		Pocock	0.38	0.19	**0.06**	**0.12**	0.56	0.36
		HP01	0.33	0.20	**0.05**	**0.13**	0.57	0.37
		HP005	0.38	0.21	**0.04**	**0.16**	0.60	0.38
	2	OBF	0.38	0.23	**0.21**	**0.18**	0.76	0.46
		Pocock	0.52	0.23	**0.06**	**0.11**	0.56	0.36
		HP01	0.45	0.23	**0.04**	**0.12**	0.57	0.36
		HP005	0.73	0.25	**0.04**	**0.15**	0.60	0.38
	3	OBF	0.81	0.27	**0.21**	**0.18**	0.76	0.46
		Pocock	0.93	0.30	**0.07**	**0.11**	0.56	0.36
		HP01	0.91	0.29	**0.05**	**0.12**	0.57	0.36
		HP005	0.91	0.30	**0.05**	**0.15**	0.60	0.38
	4	OBF	1.00	0.41	**0.22**	**0.19**	0.76	0.46
		Pocock	1.07	0.45	**0.09**	**0.13**	0.56	0.36
		HP01	1.02	0.42	**0.07**	**0.13**	0.57	0.37
		HP005	0.99	0.41	**0.07**	**0.15**	0.60	0.38
6	1	OBF	0.57	0.23	**0.29**	**0.21**	0.75	0.46
		Pocock	0.21	0.17	**0.07**	**0.11**	0.50	0.31
		HP01	0.21	0.17	**0.05**	**0.12**	0.51	0.32
		HP005	0.22	0.19	**0.04**	**0.15**	0.54	0.33
	2	OBF	0.12	0.22	**0.18**	**0.18**	0.75	0.46
		Pocock	0.16	0.20	**0.06**	**0.09**	0.50	0.31
		HP01	0.13	0.20	**0.05**	**0.10**	0.51	0.32
		HP005	0.11	0.22	**0.04**	**0.14**	0.54	0.33
	3	OBF	0.16	0.23	**0.17**	**0.17**	0.75	0.46
		Pocock	0.68	0.25	**0.07**	**0.09**	0.50	0.31
		HP01	0.52	0.24	**0.05**	**0.10**	0.51	0.32
		HP005	0.71	0.26	**0.04**	**0.13**	0.54	0.33
	4	OBF	0.86	0.28	**0.17**	**0.17**	0.75	0.46
		Pocock	0.98	0.32	**0.08**	**0.10**	0.50	0.31
		HP01	0.95	0.30	**0.06**	**0.10**	0.51	0.32
		HP005	0.94	0.31	**0.05**	**0.14**	0.54	0.33
	5	OBF	1.02	0.41	**0.19**	**0.18**	0.75	0.46
		Pocock	1.09	0.46	**0.10**	**0.12**	0.50	0.31
		HP01	1.05	0.43	**0.08**	**0.12**	0.51	0.32
		HP005	1.01	0.41	**0.07**	**0.14**	0.54	0.33

Table 5.6: Performance Score Results for Unequal Spaced 2 Times Increment Weighted Sample Size Re-estimation Designs

Total Look	Reestimation Look	Method	n_init=63		n_init=356		n_init=2018	
			R_f	LABC	R_f	LABC	R_f	LABC
3	1	OBF	0.83	0.29	**0.40**	**0.25**	0.81	0.49
		Pocock	0.57	0.26	**0.06**	**0.15**	0.62	0.42
		HP01	0.62	0.27	**0.05**	**0.18**	0.64	0.43
		HP005	0.85	0.29	**0.04**	**0.21**	0.67	0.44
	2	OBF	0.90	0.43	**0.41**	**0.25**	0.81	0.49
		Pocock	0.95	0.46	**0.06**	**0.19**	0.62	0.42
		HP01	0.91	0.44	**0.05**	**0.20**	0.64	0.43
		HP005	0.90	0.43	**0.04**	**0.22**	0.67	0.44
4	1	OBF	0.70	0.30	**0.59**	**0.27**	0.80	0.47
		Pocock	0.19	0.19	**0.07**	**0.13**	0.39	0.25
		HP01	0.18	0.20	**0.06**	**0.16**	0.41	0.26
		HP005	0.61	0.23	**0.05**	**0.20**	0.45	0.29
	2	OBF	0.83	0.29	**0.40**	**0.25**	0.80	0.47
		Pocock	0.63	0.27	**0.07**	**0.14**	0.39	0.25
		HP01	0.62	0.28	**0.05**	**0.16**	0.41	0.26
		HP005	0.86	0.29	**0.04**	**0.20**	0.45	0.29
	3	OBF	0.89	0.43	**0.41**	**0.25**	0.80	0.47
		Pocock	0.97	0.46	**0.07**	**0.18**	0.39	0.25
		HP01	0.93	0.44	**0.05**	**0.19**	0.41	0.26
		HP005	0.91	0.44	**0.04**	**0.21**	0.45	0.29
5	1	OBF	0.75	0.26	**0.62**	**0.28**	0.81	0.47
		Pocock	0.27	0.15	**0.08**	**0.13**	0.05	0.15
		HP01	0.24	0.16	**0.07**	**0.15**	0.06	0.17
		HP005	0.35	0.18	**0.06**	**0.19**	0.13	0.21
	2	OBF	0.70	0.30	**0.58**	**0.27**	0.81	0.47
		Pocock	0.20	0.20	**0.08**	**0.14**	0.05	0.15
		HP01	0.19	0.20	**0.07**	**0.15**	0.07	0.17
		HP005	0.61	0.23	**0.05**	**0.19**	0.13	0.21
	3	OBF	0.83	0.29	**0.38**	**0.25**	0.81	0.47
		Pocock	0.69	0.28	**0.07**	**0.15**	0.06	0.15
		HP01	0.63	0.28	**0.06**	**0.16**	0.07	0.17
		HP005	0.86	0.29	**0.05**	**0.19**	0.13	0.21
	4	OBF	0.90	0.43	**0.43**	**0.25**	0.81	0.47
		Pocock	0.99	0.48	**0.22**	**0.19**	0.06	0.15
		HP01	0.95	0.46	**0.06**	**0.19**	0.07	0.17
		HP005	0.92	0.44	**0.05**	**0.20**	0.13	0.21
6	1	OBF	0.75	0.26	**0.66**	**0.30**	0.81	0.47
		Pocock	0.33	0.11	**0.09**	**0.13**	0.00	0.12
		HP01	0.34	0.12	**0.08**	**0.14**	0.00	0.13
		HP005	0.29	0.14	**0.07**	**0.18**	0.00	0.18
	2	OBF	0.75	0.26	**0.64**	**0.28**	0.81	0.47
		Pocock	0.27	0.16	**0.09**	**0.14**	0.00	0.12
		HP01	0.26	0.16	**0.08**	**0.15**	0.00	0.13
		HP005	0.38	0.18	**0.06**	**0.19**	0.00	0.18
	3	OBF	0.70	0.30	**0.59**	**0.27**	0.81	0.47
		Pocock	0.24	0.21	**0.08**	**0.15**	0.00	0.12
		HP01	0.20	0.21	**0.08**	**0.15**	0.00	0.12
		HP005	0.63	0.23	**0.06**	**0.19**	0.00	0.18
	4	OBF	0.84	0.29	**0.40**	**0.25**	0.81	0.47
		Pocock	0.88	0.29	**0.08**	**0.16**	0.00	0.12

Table 5.6: Performance Score Results for Unequal Spaced 2 Times Increment Weighted Sample Size Re-estimation Designs

		HP01	0.84	0.29	**0.07**	**0.16**	0.00	0.12
		HP005	0.87	0.30	**0.05**	**0.19**	0.00	0.18
	5	OBF	0.90	0.43	**0.48**	**0.26**	0.81	0.47
		Pocock	0.99	0.48	**0.27**	**0.20**	0.00	0.12
		HP01	0.98	0.47	**0.22**	**0.19**	0.00	0.12
		HP005	0.93	0.44	**0.05**	**0.20**	0.00	0.18

Table 5.7: Performance Score Results for Equal Spaced Unweighted Sample Size Reestimation Designs

Total Looks	Reestimation Look	Method	n_{init}=356	
			R_f	ABLC
2	1	OBF	0.44	0.27
		Pocock	0.28	0.25
		HP01	0.31	0.26
		HP005	0.44	0.27
3	1	OBF	0.35	0.22
		Pocock	0.05	0.17
		HP01	0.06	0.18
		HP005	0.13	0.20
	2	OBF	0.27	0.23
		Pocock	0.03	0.18
		HP01	0.03	0.19
		HP005	0.09	0.20
4	1	OBF	0.29	0.20
		Pocock	0.04	0.13
		HP01	0.02	0.14
		HP005	0.02	0.16
	2	OBF	0.22	0.19
		Pocock	0.03	0.12
		HP01	0.01	0.13
		HP005	0.00	0.15
	3	OBF	0.21	0.21
		Pocock	0.04	0.15
		HP01	0.00	0.15
		HP005	0.00	0.17
5	1	OBF	0.26	0.20
		Pocock	0.05	0.11
		HP01	0.03	0.12
		HP005	0.02	0.15
	2	OBF	0.18	0.17
		Pocock	0.04	0.10
		HP01	0.02	0.11
		HP005	0.01	0.13
	3	OBF	0.16	0.17
		Pocock	0.04	0.10
		HP01	0.01	0.11
		HP005	0.00	0.13
	4	OBF	0.16	0.19
		Pocock	0.04	0.13
		HP01	0.01	0.13
		HP005	0.00	0.15

Table 5.8: Performance Score Results for Unequal Spaced 2-times Increment Weighted Sample Size Re-estimation Designs

Total Looks	Reestimation Look	Method	n_{init}=356	
			R_f	ABLC
3	1	OBF	0.36	0.23
		Pocock	0.04	0.14
		HP01	0.02	0.16
		HP005	0.02	0.19
	2	OBF	0.48	0.26
		Pocock	0.03	0.19
		HP01	0.01	0.20
		HP005	0.00	0.22
4	1	OBF	0.52	0.25
		Pocock	0.06	0.11
		HP01	0.05	0.13
		HP005	0.04	0.17
	2	OBF	0.36	0.23
		Pocock	0.05	0.12
		HP01	0.03	0.14
		HP005	0.03	0.18
	3	OBF	0.38	0.25
		Pocock	0.04	0.18
		HP01	0.02	0.18
		HP005	0.00	0.21
5	1	OBF	0.58	0.27
		Pocock	0.08	0.10
		HP01	0.07	0.12
		HP005	0.05	0.16
	2	OBF	0.53	0.25
		Pocock	0.07	0.11
		HP01	0.06	0.12
		HP005	0.04	0.16
	3	OBF	0.36	0.22
		Pocock	0.06	0.13
		HP01	0.04	0.14
		HP005	0.03	0.17
	4	OBF	0.41	0.25
		Pocock	0.19	0.19
		HP01	0.03	0.18
		HP005	0.01	0.20

5.3.2 Comparison of performance

A. Total number of looks

For group sequential designs, when the maximum sample size is small, performance is improved by increasing the number of total looks. As show in Tables 5.3 and 5.4, when maximum sample size is 63 or 356, both the failure rate and the log area between curves are similar across designs with different total number of looks. However, when the maximum sample size is large, the performance improved dramatically when total number of looks is increased.

B. Type of sample size increment

Performance for adaptive designs with different patient increments are evaluated. For GS designs with a maximum sample size of 2018, except designs using O'Brien and Fleming boundaries, performance is improved when 2-times increment type is selected (Table 5.9). Both the failure rate and log area between curves are decreased. Performance improvement get larger when the total number of looks increases. However, because the total sample sizes are very small for GS designs with maximum sample size 63 and 356, no apparent performance difference is observed (see Tables 5.3 and 5.4). There is no performance improvement observed in weighted and unweighted sample size re-estimation designs, regardless total number of looks and type of boundaries as in Tables 5.5 to 5.8.

Table 5.9: Comparison of Group Sequential Designs with Different Patient Increments

		equal space n_{max}=2018		2-times increment n_{max}=2018	
Total Looks	Method	R_f	ABLC	R_f	ABLC
3	OBF	0.81	0.54	0.81	0.49
	Pocock	0.70	0.51	0.62	0.42
	HP01	0.72	0.52	0.64	0.43
	HP005	0.74	0.52	0.67	0.44
4	OBF	0.79	0.49	0.81	0.47
	Pocock	0.62	0.42	0.39	0.25
	HP01	0.64	0.43	0.40	0.27
	HP005	0.66	0.44	0.44	0.29
5	OBF	0.76	0.46	0.81	0.47
	Pocock	0.56	0.36	0.05	0.15
	HP01	0.57	0.37	0.07	0.17
	HP005	0.61	0.38	0.14	0.21
6	OBF	0.75	0.46	0.81	0.47
	Pocock	0.50	0.32	0.00	0.12
	HP01	0.51	0.32	0.00	0.13
	HP005	0.54	0.33	0.00	0.18

C. Comparison of two-stage adaptive designs

Two-stage adaptive designs have one interim analysis and one final analysis. Because of the simplicity of two-stage designs, it is most often used in clinical trials. Comparison of performance of GS design with maximum sample size of 2018 and weighted and unweighted re-estimation designs with initial sample size 356 is presented in Table 5.10. As there is only one interim analysis before the final look, the average sample size for GS design is at least 1009 and it is over-sized for a large portion of the treatment effect intervals. Weighted or unweighted re-estimation designs start with a much smaller initial sample size and sample size will only be increased when the interim finding indicates a small treatment effect. Thus, the performance of two-stage re-estimation designs is better

than the two-stage group sequential designs. The performance of weighted re-estimation designs is almost identical to the performance of unweighted re-estimation designs.

Table 5.10: Comparison of Two-stage Adaptive Designs (n_{max}=2018)

	GS		Weighted n_{init}=356		Unweighted n_{init}=356	
Method	R_f	ABLC	R_f	ABLC	R_f	ABLC
OBF	0.82	0.65	0.44	0.26	0.44	0.27
Pocock	0.79	0.64	0.31	0.25	0.28	0.25
HP01	0.81	0.65	0.36	0.26	0.31	0.26
HP005	0.82	0.65	0.51	0.27	0.44	0.27

D. Comparison of two-stage group sequential designs

A two-stage design has an interim analysis at the end of stage one and the sample size in stage two can be modified. In case a sequential design is introduced in stage two, it becomes a two-stage group sequential design (See Figure 5.1). This design is equivalent to the sample size re-estimation design with sample size re-estimation at the first look. Comparison of two-stage GS designs are presented in Table 5.11 and 5.12. In general, failure rate for unweighted designs are slightly smaller than the weighted designs, while ABLC for these two types of designs are almost identical.

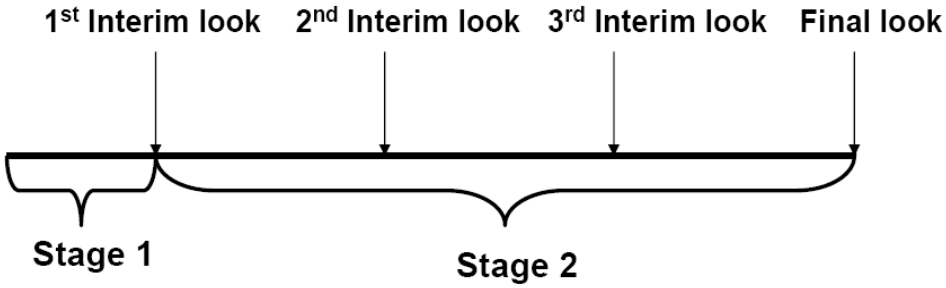

Figure 5.1: Illustration of Two-stage Group Sequential Design

Table 5.11: Comparison of Equal-spaced Two-stage Group Sequential Designs

Total Looks	Method	Weighted n_{init}=356		Unweighted n_{init}=356	
		R_f	ABLC	R_f	ABLC
2	OBF	0.44	0.26	0.44	0.27
	Pocock	0.31	0.25	0.28	0.25
	HP01	0.36	0.26	0.31	0.26
	HP005	0.51	0.27	0.44	0.27
3	OBF	0.39	0.23	0.35	0.22
	Pocock	0.07	0.17	0.05	0.17
	HP01	0.10	0.19	0.06	0.18
	HP005	0.17	0.21	0.13	0.20
4	OBF	0.32	0.21	0.29	0.20
	Pocock	0.06	0.14	0.04	0.13
	HP01	0.05	0.15	0.02	0.14
	HP005	0.04	0.18	0.02	0.16
5	OBF	0.28	0.21	0.26	0.20
	Pocock	0.06	0.12	0.05	0.11
	HP01	0.05	0.13	0.03	0.12
	HP005	0.04	0.16	0.02	0.15

Table 5.12: Comparison of Unequal-spaced Two-stage Group Sequential Designs

Total Looks	Method	Weighted n_{init}=356		Unweighted n_{init}=356	
		R_f	ABLC	R_f	ABLC
3	OBF	0.40	0.25	0.36	0.23
	Pocock	0.06	0.15	0.04	0.14
	HP01	0.05	0.18	0.02	0.16
	HP005	0.04	0.21	0.02	0.19
4	OBF	0.59	0.27	0.52	0.25
	Pocock	0.07	0.13	0.06	0.11
	HP01	0.06	0.16	0.05	0.13
	HP005	0.05	0.20	0.04	0.17
5	OBF	0.62	0.28	0.58	0.27
	Pocock	0.08	0.13	0.08	0.10
	HP01	0.07	0.15	0.07	0.12
	HP005	0.06	0.19	0.05	0.16

E. Comparison of 5-look unequal-spaced GS designs using HP type boundaries versus sample size re-estimation designs using O'Brien and Fleming boundaries and Pocock type boundaries

Since sample size can be adjusted based on interim findings, failure rate and ABLC for sample size re-estimation designs are usually low. However in practice, there is still a lack of understanding on sample size re-estimation designs compared to group sequential designs. Thus, sample size re-estimation designs are not well accepted by regulatory agencies (FDA, etc.). By comparing unequal-spaced (2-times increment) 5-look GS designs using HP01 and HP005 boundaries with sample size re-estimation designs using OBF and PK boundaries, it can be seen that failure rate and ABLC for GS designs are lower than or similar to those of re-estimation designs as shown in Tables 5.13 and 5.14 and Figure 5.2 and 5.9.

Table 5.13: Comparison of 5-look 2-times increment GS design with HP01 boundaries versus equal-spaced sample size re-estimation designs with OBF and PK boundaries

Total Looks	Reestimation Look	method	HP01-Weighted		HP01-Unweighted	
			Difference of R_f	Difference of ABLC	Difference of R_f	Difference of ABLC
2	1	OBF	-0.3646	-0.0970	-0.3650	-0.0981
		Pocock	-0.2420	-0.0836	-0.2102	-0.0840
3	1	OBF	-0.3144	-0.0580	-0.2765	-0.0529
		Pocock	0.0023	-0.0041	0.0202	-0.0004
	2	OBF	-0.2463	-0.0610	-0.2007	-0.0613
		Pocock	0.0041	-0.0140	0.0423	-0.0125
4	1	OBF	-0.2471	-0.0458	-0.2210	-0.0358
		Pocock	0.0133	0.0315	0.0280	0.0344
	2	OBF	-0.1956	-0.0274	-0.1511	-0.0210
		Pocock	0.0097	0.0416	0.0357	0.0476
	3	OBF	-0.1982	-0.0402	-0.1389	-0.0407
		Pocock	-0.0103	0.0179	0.0346	0.0194
5	1	OBF	-0.2127	-0.0416	-0.1869	-0.0277
		Pocock	0.0100	0.0471	0.0221	0.0539
	2	OBF	-0.1404	-0.0160	-0.1110	-0.0054
		Pocock	0.0097	0.0621	0.0309	0.0724
	3	OBF	-0.1365	-0.0090	-0.0871	-0.0010
		Pocock	0.0021	0.0590	0.0325	0.0720
	4	OBF	-0.1471	-0.0255	-0.0876	-0.0248
		Pocock	-0.0213	0.0331	0.0297	0.0340

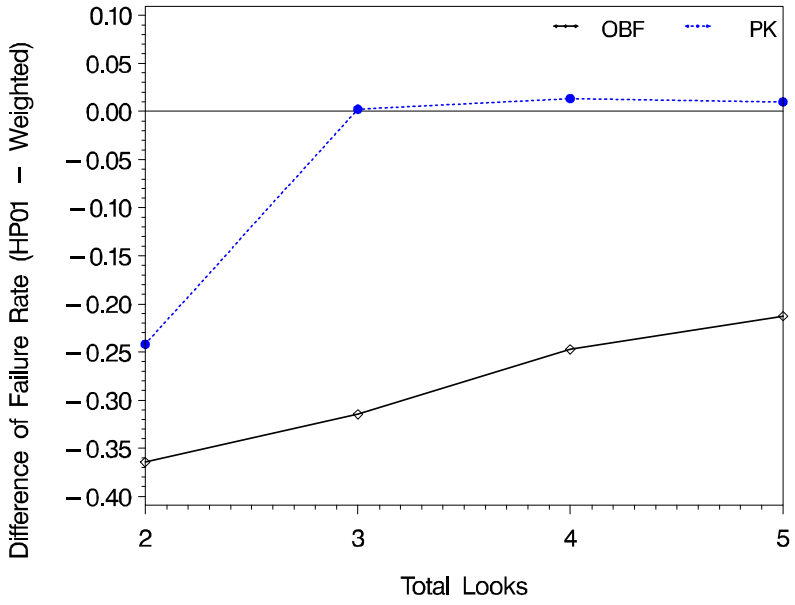

Figure 5.2: Failure Rate of 5-look Unequal-spaced GS Design using HP01 Boundary versus Weighted Designs using OBF & PK Boundary

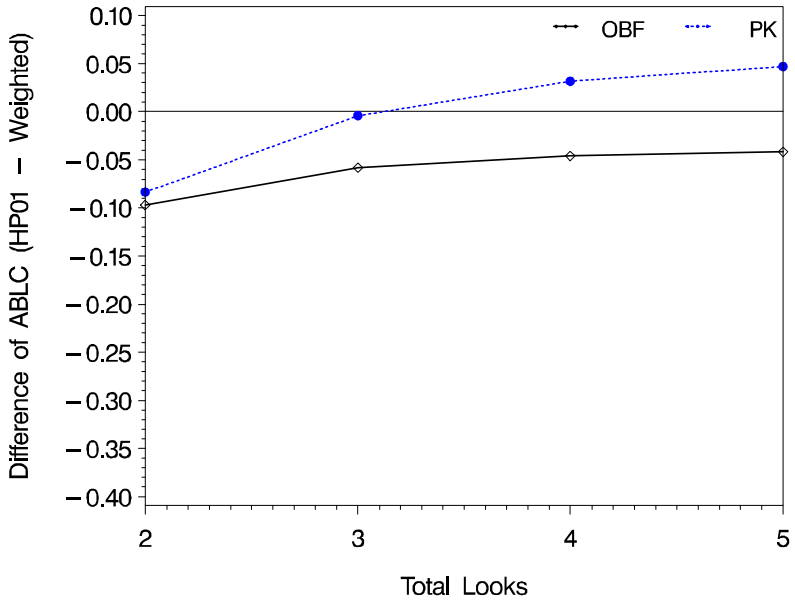

Figure 5.3: ABLC of 5-look Unequal-spaced GS Design using HP01 Boundary versus Weighted Designs using OBF & PK Boundary

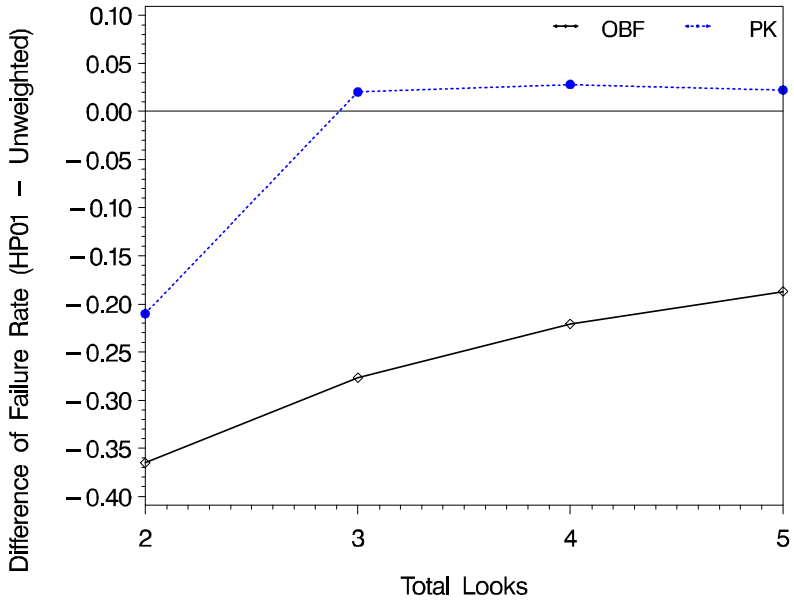

Figure 5.4: Failure Rate of 5-look Unequal-spaced GS Design using HP01 Boundary versus Unweighted Designs using OBF & PK Boundary

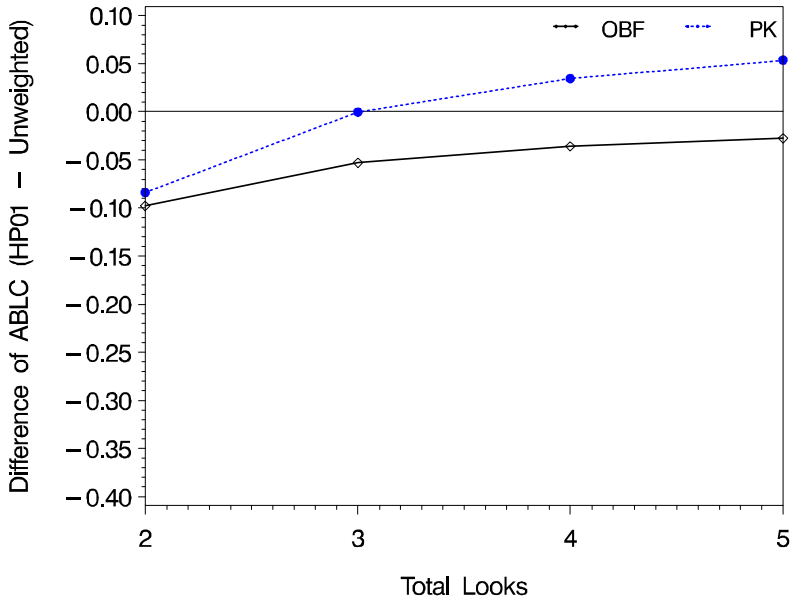

Figure 5.5: ABLC of 5-look Unequal-spaced GS Design using HP01 Boundary versus Unweighted Designs using OBF & PK Boundary

Table 5.14: Comparison of 5-look 2-times increment GS design with HP005 boundaries versus equal-spaced sample size re-estimation designs with OBF and PK boundaries

Total Looks	Reestimation Look	method	HP005-Weighted		HP005-Unweighted	
			Difference of R_f	Difference of ABLC	Difference of R_f	Difference of ABLC
2	1	OBF	-0.2981	-0.0568	-0.2985	-0.0579
		Pocock	-0.1755	-0.0434	-0.1437	-0.0438
3	1	OBF	-0.2479	-0.0178	-0.2100	-0.0127
		Pocock	0.0688	0.0361	0.0867	0.0398
	2	OBF	-0.1798	-0.0208	-0.1342	-0.0211
		Pocock	0.0706	0.0262	0.1088	0.0277
4	1	OBF	-0.1806	-0.0056	-0.1545	0.0044
		Pocock	0.0798	0.0717	0.0945	0.0746
	2	OBF	-0.1291	0.0128	-0.0846	0.0192
		Pocock	0.0762	0.0818	0.1022	0.0878
	3	OBF	-0.1317	<0.0001	-0.0724	-0.0005
		Pocock	0.0562	0.0581	0.1011	0.0596
5	1	OBF	-0.1462	-0.0014	-0.1204	0.0125
		Pocock	0.0765	0.0873	0.0886	0.0941
	2	OBF	-0.0739	0.0242	-0.0445	0.0348
		Pocock	0.0762	0.1023	0.0974	0.1126
	3	OBF	-0.0700	0.0312	-0.0206	0.0392
		Pocock	0.0686	0.0992	0.0990	0.1122
	4	OBF	-0.0806	0.0147	-0.0211	0.0154
		Pocock	0.0452	0.0733	0.0962	0.0742

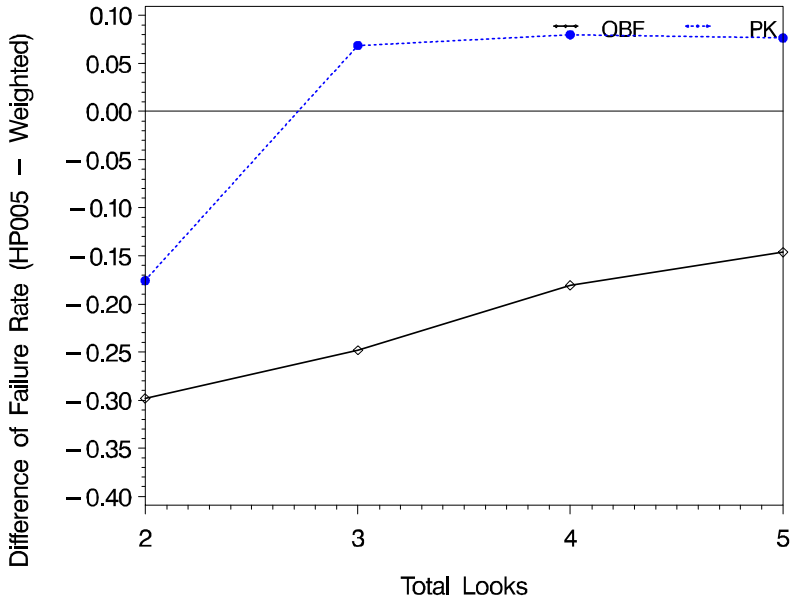

Figure 5.6: Failure Rate of 5-look Unequal-spaced GS Design using HP005 Boundary versus Weighted Designs using OBF & PK Boundary

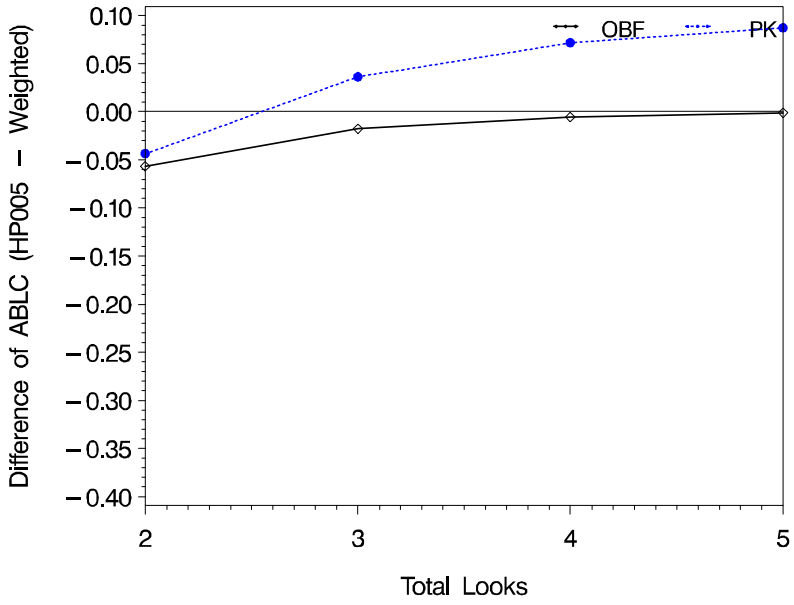

Figure 5.7: ABLC of 5-look Unequal-spaced GS Design using HP005 Boundary versus Weighted Designs using OBF & PK Boundary

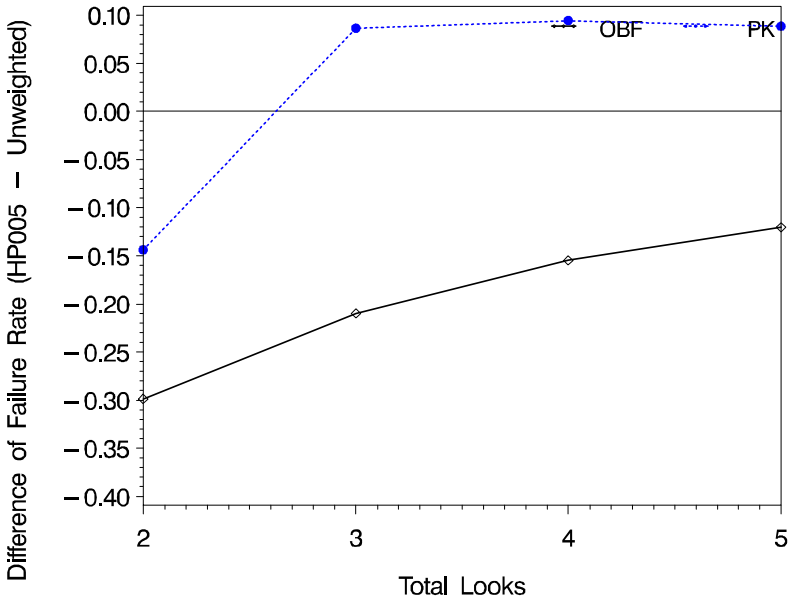

Figure 5.8: Failure Rate of 5-look Unequal-spaced GS Design using HP005 Boundary versus Unweighted Designs using OBF & PK Boundary

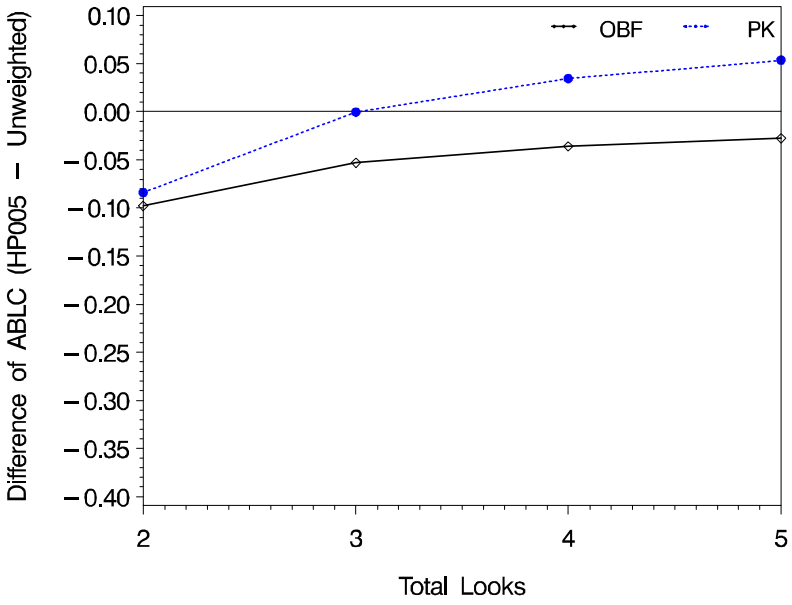

Figure 5.9: ABLC of 5-look Unequal-spaced GS Design using HP005 Boundary versus Unweighted Designs using OBF & PK Boundary

F. Comparison of performance under different treatment effect intervals

Is adaptive design always the best choice? As shown in Tables 5.15 to 5.18, performance of the adaptive design is also affected by the location and length of the treatment effect interval. For some of the intervals, both the failure rate and ABLC can be very low. As indicated in Figure 5.10, when the adaptive index is small, performance on the treatment effect interval could be very robust. In this case, regardless what treatment effect on the interval we chose to design a fix sample size study, the performance for fixed sample size design and adaptive designs will be very close to each other. Thus, we can just either pick the lower limit or the middle point of the treatment effect interval to design a fixed sample size study. However, in order to obtain such a treatment effect interval, we have to have sufficient data to support the determination of this interval. Thus, it might be quite difficult to find such an accurate estimate of interval. And in terms of adaptive index, how small is small? So even if a small interval can be obtained, simulations may still need to be done to compare the performance before making a decision to use fixed sample size design rather than adaptive designs.

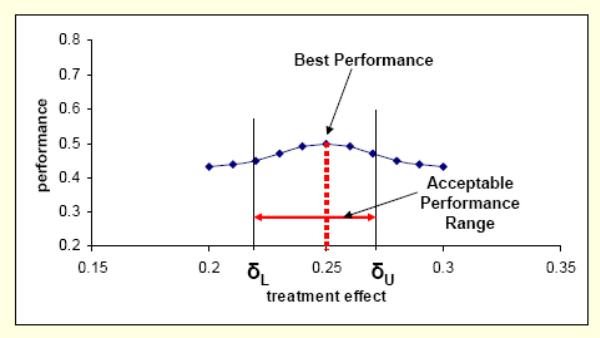

Figure 5.10: Illustration of performance on a narrow treatment effect interval

Table 5.15: Performance for Equal Spaced GS Designs Under Different Treatment Effect Intervals

Total Looks	Method	Treatment Effect Interval							
		(0.0882,0.2)		(0.2,0.3)		(0.3,0.4)		(0.4,0.5)	
		R_f	ABLC	R_f	ABLC	R_f	ABLC	R_f	ABLC
2	OBF	0.7782	0.1048	<0.0001	0.0184	0.1754	0.0579	1.0000	0.0898
	Pocock	0.8746	0.1272	<0.0001	0.0165	<0.0001	0.0439	0.8736	0.0843
	HP01	0.7910	0.1100	<0.0001	0.0165	0.0216	0.0524	1.0000	0.0873
	HP005	0.7783	0.1047	<0.0001	0.0186	0.1845	0.0581	1.0000	0.0899
3	OBF	0.7836	0.1127	<0.0001	0.0168	0.0129	0.0527	1.0000	0.0822
	Pocock	0.9109	0.1394	<0.0001	0.0215	<0.0001	0.0236	<0.0001	0.0538
	HP01	0.8017	0.1187	<0.0001	0.0146	<0.0001	0.0342	0.1155	0.0598
	HP005	0.7843	0.1105	<0.0001	0.0154	<0.0001	0.0429	0.3218	0.0659
4	OBF	0.7850	0.1171	<0.0001	0.0157	<0.0001	0.0467	0.8423	0.0789
	Pocock	0.9371	0.1451	<0.0001	0.0263	<0.0001	0.0147	<0.0001	0.0389
	HP01	0.8230	0.1252	<0.0001	0.0163	<0.0001	0.0247	<0.0001	0.0457
	HP005	0.7862	0.1139	<0.0001	0.0140	<0.0001	0.0342	<0.0001	0.0540
5	OBF	0.8015	0.1196	<0.0001	0.0156	<0.0001	0.0423	0.6274	0.0728
	Pocock	0.9657	0.1504	<0.0001	0.0302	<0.0001	0.0095	<0.0001	0.0306
	HP01	0.8514	0.1301	<0.0001	0.0191	<0.0001	0.0176	<0.0001	0.0366
	HP005	0.8084	0.1173	<0.0001	0.0142	<0.0001	0.0289	<0.0001	0.0458

Table 5.16: Performance for Unequal Spaced GS Designs Under Different Treatment Effect Intervals

Total Looks	Method	Treatment Effect Interval							
		(0.0882,0.2)		(0.2,0.3)		(0.3,0.4)		(0.4,0.5)	
		R_f	ABLC	R_f	ABLC	R_f	ABLC	R_f	ABLC
3	OBF	0.7690	0.1053	<0.0001	0.0183	0.0314	0.0558	1.0000	0.0815
	Pocock	0.9431	0.1382	<0.0001	0.0202	<0.0001	0.0185	<0.0001	0.0388
	HP01	0.8251	0.1181	<0.0001	0.0145	<0.0001	0.0322	<0.0001	0.0484
	HP005	0.7858	0.1096	<0.0001	0.0159	<0.0001	0.0431	<0.0001	0.0576
4	OBF	0.7864	0.1054	<0.0001	0.0179	0.0256	0.0558	1.0000	0.0815
	Pocock	0.9862	0.1438	<0.0001	0.0210	<0.0001	0.0143	<0.0001	0.0250
	HP01	0.8730	0.1239	<0.0001	0.0145	<0.0001	0.0259	<0.0001	0.0343
	HP005	0.8097	0.1118	<0.0001	0.0147	<0.0001	0.0391	<0.0001	0.0481
5	OBF	0.7694	0.1053	<0.0001	0.0181	0.0172	0.0555	1.0000	0.0812
	Pocock	1.0000	0.1466	0.0379	0.0209	<0.0001	0.0151	<0.0001	0.0234
	HP01	0.9270	0.1314	<0.0001	0.0154	<0.0001	0.0221	<0.0001	0.0301
	HP005	0.8356	0.1144	<0.0001	0.0149	<0.0001	0.0372	<0.0001	0.0452

Table 5.17: Performance for Equal Spaced Weighted Designs Under Different Treatment Effect Intervals

Total Looks	Method	Treatment Effect Interval							
		(0.0882,0.2)		(0.2,0.3)		(0.3,0.4)		(0.4,0.5)	
		R_f	ABLC	R_f	ABLC	R_f	ABLC	R_f	ABLC
2	OBF	0.1417	0.0357	<0.0001	0.0653	0.6337	0.0720	1.0000	0.0916
	Pocock	0.1799	0.0371	<0.0001	0.0637	0.0863	0.0634	1.0000	0.0869
	HP01	0.1536	0.0355	<0.0001	0.0645	0.2943	0.0681	1.0000	0.0891
	HP005	0.1444	0.0363	<0.0001	0.0662	0.9225	0.0725	1.0000	0.0916
3	OBF	0.1362	0.0248	<0.0001	0.0466	0.4327	0.0681	1.0000	0.0861
	Pocock	0.1912	0.0227	<0.0001	0.0397	<0.0001	0.0490	0.0676	0.0603
	HP01	0.1585	0.0255	<0.0001	0.0431	<0.0001	0.0533	0.2365	0.0649
	HP005	0.1440	0.0286	<0.0001	0.0482	<0.0001	0.0584	0.5554	0.0705
4	OBF	0.1421	0.0241	<0.0001	0.0430	0.1499	0.0621	1.0000	0.0842
	Pocock	0.2119	0.0188	<0.0001	0.0292	<0.0001	0.0394	<0.0001	0.0487
	HP01	0.1692	0.0218	<0.0001	0.0334	<0.0001	0.0443	<0.0001	0.0526
	HP005	0.1577	0.0254	<0.0001	0.0418	<0.0001	0.0507	<0.0001	0.0595
5	OBF	0.1570	0.0237	<0.0001	0.0446	<0.0001	0.0618	0.9916	0.0791
	Pocock	0.2233	0.0193	<0.0001	0.0238	<0.0001	0.0341	<0.0001	0.0433
	HP01	0.1764	0.0203	<0.0001	0.0285	<0.0001	0.0370	<0.0001	0.0457
	HP005	0.1547	0.0230	<0.0001	0.0369	<0.0001	0.0462	<0.0001	0.0536

Table 5.18: Performance for Unequal Spaced Weighted Designs Under Different Treatment Effect Intervals

Total Looks	Method	Treatment Effect Interval							
		(0.0882,0.2)		(0.2,0.3)		(0.3,0.4)		(0.4,0.5)	
		R_f	ABLC	R_f	ABLC	R_f	ABLC	R_f	ABLC
3	OBF	0.1300	0.0350	<0.0001	0.0588	0.4847	0.0692	1.0000	0.0847
	Pocock	0.2164	0.0274	<0.0001	0.0373	<0.0001	0.0396	<0.0001	0.0481
	HP01	0.1679	0.0316	<0.0001	0.0471	<0.0001	0.0490	<0.0001	0.0547
	HP005	0.1452	0.0349	<0.0001	0.0562	<0.0001	0.0587	<0.0001	0.0630
4	OBF	0.1596	0.0305	0.2714	0.0617	1.0000	0.0861	1.0000	0.0961
	Pocock	0.2571	0.0226	<0.0001	0.0365	<0.0001	0.0378	<0.0001	0.0360
	HP01	0.2042	0.0254	<0.0001	0.0422	<0.0001	0.0469	<0.0001	0.0440
	HP005	0.1722	0.0292	<0.0001	0.0518	<0.0001	0.0606	<0.0001	0.0563
5	OBF	0.1788	0.0294	0.3736	0.0646	1.0000	0.0883	1.0000	0.1006
	Pocock	0.3097	0.0208	<0.0001	0.0304	<0.0001	0.0404	<0.0001	0.0404
	HP01	0.2541	0.0220	<0.0001	0.0335	<0.0001	0.0449	<0.0001	0.0465
	HP005	0.2195	0.0261	<0.0001	0.0451	<0.0001	0.0574	<0.0001	0.0602

Chapter 6

Performance of Adaptive Designs when Treatment Effect Follows a Beta Distribution

6.1 Method

In the previous chapter, performance of the adaptive designs were evaluated when considered treatment effect follows a uniform distribution on an treatment effect interval. However, some treatment effects may be more likely observed in practice. So it is logical to consider treatment effect as a random variable on the treatment effect interval. In this chapter, it is assumed that the treatment effect follows a beta distribution.

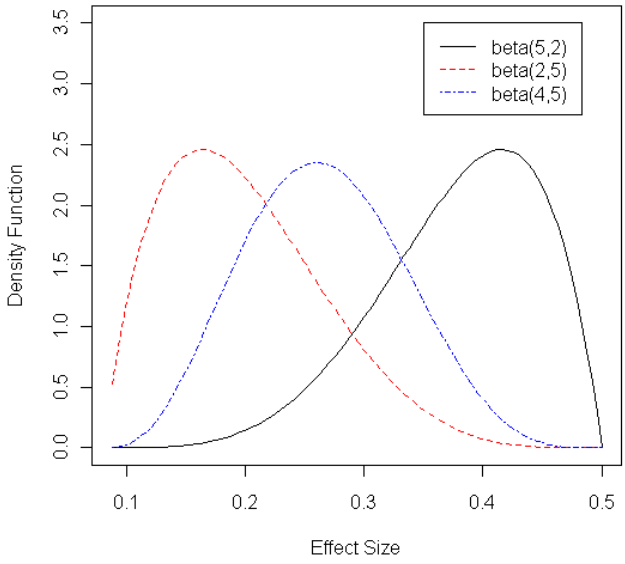

Figure 6.1: Beta distributions for treatment effect

6.2 Simulation Plan

The same simulation procedures and parameters in the previous chapter were used. Treatment effects were randomly generated and assumed follows a beta distribution - $Beta\,(5,2)$, $Beta\,(2,5)$, or $Beta\,(4,5)$ (see Figure 6.1). For each randomly generated treatment effect, sample size and result (whether accept the null hypothesis or the alternative hypothesis) were obtained from the simulation study. Based on grouped treatment effects, sample size curve and power curve were obtained by interpolation. Performance scores were calculated by comparing size or power curve with the curve for true treatment effect function. 10,000 treatment effects are random generated for each simulation study.

6.3 Results

6.3.1 Performance of adaptive designs

Performance for each adaptive design on the treatment effect interval [0.0882, 0.5] was measured when treatment effect follows a beta distribution. For group sequential design, performance was evaluated when maximum allowed sample size is 2018. Performance of sample size re-estimation designs were evaluated with the initial sample size is 356 and the sample size can be adjusted to maximum 2018 after sample size re-estimation.. Performance scores for equal-spaced (patient increment) group sequential designs, weighted sample size re-estimation designs, and unweighted sample size re-estimation designs are presented in Tables 6.1, 6.3, and 6.5, respectively. Performance for designs with unequal-spaced (2-times) patient increment are displayed in Tables 6.2, 6.4, and 6.6 respectively. Sample size and power curves for selected adaptive designs when treatment effect follows a beta distribution are displayed in Appendix C. Performance for designs with treatment effect follows a distribution of beta(5, 2) is much better than the performance for designs with treatment effect follows a beta(2, 5) or bets(4, 5).

Table 6.1: Performance Score Results for Equal Spaced GS Designs

		Beta(5, 2)		Beta(2, 5)		Beta(4, 5)	
Total Looks	Method	R_f	ABLC	R_f	ABLC	R_f	ABLC
2	OBF	0.7812	0.6458	0.8096	0.6529	0.8301	0.6597
	Pocock	0.7812	0.6466	0.7894	0.6416	0.8035	0.6487
	HP01	0.7812	0.6460	0.8004	0.6490	0.8155	0.6538
	HP005	0.7812	0.6460	0.8092	0.6522	0.8280	0.6603
5	OBF	0.5680	0.3765	0.7592	0.5146	0.7894	0.5242
	Pocock	0.5310	0.3934	0.6209	0.3705	0.6216	0.3929
	HP01	0.5315	0.3923	0.6239	0.3809	0.6535	0.4009
	HP005	0.5320	0.3877	0.6531	0.3936	0.7125	0.4255

Table 6.2: Performance Score Results for Unequal Spaced GS Designs

Total Looks	Method	Beta(5, 2)		Beta(2, 5)		Beta(4, 5)	
		R_f	ABLC	R_f	ABLC	R_f	ABLC
5	OBF	0.5848	0.3813	0.7890	0.5205	0.8160	0.5336
	Pocock	0.0469	0.2519	0.3705	0.2205	0.5382	0.2564
	HP01	0.0628	0.2502	0.5121	0.2447	0.5628	0.2720
	HP005	0.0943	0.2362	0.6251	0.2891	0.7294	0.3225

Table 6.3: Performance Score Results for Equal Spaced Weighted Sample Size Reestimation Designs

Total Looks	Reestimation Look	Method	Beta(5, 2)		Beta(2, 5)		Beta(4, 5)	
			R_f	ABLC	R_f	ABLC	R_f	ABLC
2	1	OBF	0.1602	0.2426	0.6353	0.3256	0.7382	0.3787
		Pocock	<0.0001	0.2426	0.6387	0.3140	0.7584	0.3748
		HP01	<0.0001	0.2375	0.6066	0.3217	0.7538	0.3655
		HP005	<0.0001	0.2261	0.6688	0.3249	0.7362	0.3747
5	1	OBF	0.1275	0.2467	0.5433	0.2766	0.5910	0.3017
		Pocock	<0.0001	0.2580	0.2684	0.2051	0.3392	0.2406
		HP01	<0.0001	0.2453	0.3594	0.2141	0.4533	0.2453
		HP005	<0.0001	0.2331	0.4352	0.2418	0.5288	0.2752
	2	OBF	0.1233	0.2505	0.4280	0.2374	0.5484	0.2829
		Pocock	<0.0001	0.2608	0.1777	0.1830	0.3504	0.2275
		HP01	<0.0001	0.2453	0.2779	0.1941	0.3967	0.2435
		HP005	<0.0001	0.2331	0.4676	0.2279	0.5072	0.2716
	3	OBF	0.1233	0.2505	0.4192	0.2313	0.5562	0.2778
		Pocock	<0.0001	0.2608	0.0570	0.1690	0.4123	0.2384
		HP01	<0.0001	0.2537	0.2064	0.1838	0.4394	0.2623
		HP005	<0.0001	0.2401	0.3660	0.2107	0.5489	0.2683
	4	OBF	0.1210	0.2511	0.4590	0.2408	0.6734	0.3002
		Pocock	<0.0001	0.2542	0.2705	0.2118	0.5389	0.2831
		HP01	<0.0001	0.2488	0.3119	0.2063	0.4948	0.2681
		HP005	<0.0001	0.2417	0.4635	0.2137	0.5353	0.2845

Table 6.4: Performance Score Results for Unequal Spaced Weighted Sample Size Reestimation Designs

Total Looks	Reestimation Look	Method	Beta(5, 2)		Beta(2, 5)		Beta(4, 5)	
			R_f	ABLC	R_f	ABLC	R_f	ABLC
5	1	OBF	0.2654	0.2428	0.6327	0.3543	0.6530	0.3816
		Pocock	<0.0001	0.2311	0.3879	0.2299	0.3901	0.2392
		HP01	<0.0001	0.2222	0.4515	0.2398	0.5201	0.2669
		HP005	<0.0001	0.2095	0.5252	0.2749	0.6442	0.2977
	2	OBF	0.2530	0.2396	0.6019	0.3278	0.7453	0.3627
		Pocock	<0.0001	0.2271	0.4931	0.2333	0.4343	0.2721
		HP01	<0.0001	0.2141	0.4548	0.2402	0.4817	0.2872
		HP005	<0.0001	0.2066	0.5523	0.2728	0.6965	0.3265
	3	OBF	0.1875	0.2389	0.5741	0.3054	0.7230	0.3528
		Pocock	<0.0001	0.2248	0.4478	0.2295	0.6385	0.2828
		HP01	<0.0001	0.2133	0.4478	0.2415	0.5550	0.2959
		HP005	<0.0001	0.2031	0.5647	0.2908	0.7041	0.3303
	4	OBF	0.1620	0.2271	0.6388	0.3253	0.7564	0.3733
		Pocock	<0.0001	0.2087	0.7011	0.3230	0.6807	0.3461
		HP01	<0.0001	0.2033	0.6259	0.2875	0.6230	0.3296
		HP005	<0.0001	0.1935	0.6147	0.3012	0.7181	0.3510

Table 6.5: Performance Score Results for Equal Spaced Unweighted Sample Size Reestimation Designs

Total Looks	Reestimation Look	Method	Beta(5, 2)		Beta(2, 5)		Beta(4, 5)	
			R_f	ABLC	R_f	ABLC	R_f	ABLC
2	1	OBF	0.2212	0.2404	0.6361	0.3287	0.7378	0.3656
		Pocock	0.2023	0.2484	0.6245	0.3065	0.7525	0.3622
		HP01	0.2161	0.2470	0.6391	0.3188	0.7394	0.3712
		HP005	0.2195	0.2437	0.6352	0.3268	0.7359	0.3649
5	1	OBF	0.1337	0.2494	0.4656	0.2602	0.5672	0.2863
		Pocock	<0.0001	0.2473	0.3132	0.2094	0.2854	0.2174
		HP01	<0.0001	0.2470	0.3040	0.2157	0.3258	0.2318
		HP005	<0.0001	0.2352	0.3883	0.2378	0.5063	0.2586
	2	OBF	0.1208	0.2531	0.3770	0.2380	0.5288	0.2674
		Pocock	<0.0001	0.2635	0.2415	0.1875	0.2748	0.2177
		HP01	<0.0001	0.2550	0.2020	0.1968	0.2740	0.2219
		HP005	<0.0001	0.2372	0.3333	0.2135	0.5067	0.2562
	3	OBF	0.1277	0.2573	0.3701	0.2322	0.5201	0.2675
		Pocock	<0.0001	0.2624	0.1011	0.1799	0.3136	0.2254
		HP01	<0.0001	0.2573	0.2619	0.1901	0.3323	0.2296
		HP005	<0.0001	0.2448	0.3535	0.2043	0.4187	0.2447
	4	OBF	0.1072	0.2490	0.3930	0.2324	0.7007	0.3035
		Pocock	<0.0001	0.2560	0.2696	0.2100	0.5552	0.2814
		HP01	<0.0001	0.2469	0.3333	0.2111	0.5190	0.2659
		HP005	<0.0001	0.2374	0.3524	0.2182	0.5329	0.2808

Table 6.6: Performance Score Results for Unequal Spaced Unweighted Sample Size Reestimation Designs

Total Looks	Reestimation Look	Method	Beta(5, 2)		Beta(2, 5)		Beta(4, 5)	
			R_f	ABLC	R_f	ABLC	R_f	ABLC
5	1	OBF	0.2514	0.2452	0.6163	0.3339	0.6730	0.3560
		Pocock	<0.0001	0.2518	0.2987	0.2048	0.3511	0.2128
		HP01	<0.0001	0.2384	0.3474	0.2140	0.3839	0.2333
		HP005	<0.0001	0.2221	0.4676	0.2405	0.5410	0.2744
	2	OBF	0.2180	0.2333	0.5571	0.3088	0.6306	0.3365
		Pocock	<0.0001	0.2437	0.3688	0.2063	0.3299	0.2355
		HP01	<0.0001	0.2358	0.3500	0.2140	0.4340	0.2492
		HP005	<0.0001	0.2230	0.4603	0.2432	0.5668	0.2822
	3	OBF	0.1778	0.2432	0.5282	0.2812	0.6070	0.3149
		Pocock	<0.0001	0.2338	0.4804	0.2269	0.4452	0.2488
		HP01	<0.0001	0.2252	0.4597	0.2348	0.4532	0.2614
		HP005	<0.0001	0.2109	0.5014	0.2570	0.5735	0.2905
	4	OBF	0.1707	0.2245	0.6009	0.3166	0.7343	0.3689
		Pocock	<0.0001	0.2061	0.6566	0.3002	0.7129	0.3562
		HP01	<0.0001	0.2063	0.6465	0.2777	0.6509	0.3369
		HP005	<0.0001	0.1983	0.6182	0.2949	0.6381	0.3325

6.3.2 Comparison of performance

Comparison of 5-look unequal-spaced GS designs using HP type boundaries versus sample size re-estimation designs using O'Brien and Fleming boundaries and Pocock type boundaries are presented in Tables 6.7 to 6.12 and Figures 6.2 to 6.25. Similar to the comparisons when treatment effect follows a uniform distribution, performance for 5-look unequal spaced GS designs are consistently comparable to the performance of equal-spaced sample size re-estimation designs with OBF and PK boundaries. However, the two performance scores may not always show the same magnitude of improvement. When sample size or power for those treatment effects is substantially deviated from the true treatment effect function, it could happen that failure rate shows improvement while ABLC doesn't. On the other hand, when sample size or power for those treatment effects is only slightly deviated from true treatment effect function, it could happen that ABLC shows improvement while failure rate doesn't.

A) Treatment effect follows Beta(5,2)

Table 6.7: Comparison of 5-look 2-times increment random GS design with HP01 boundaries versus equal-spaced random sample size re-estiamtion designs with OBF and PK boundaries when treatment effect follows Beta(5,2)

Total Looks	Reestimation Look	Method	HP01-Weighted		HP01-Unweighted	
			Difference of R_f	Difference of ABLC	difference of R_f	Difference of ABLC
2	1	OBF	-0.1599	0.0038	-0.1584	0.0098
		Pocock	-0.1402	0.0039	-0.1395	0.0018
5	1	OBF	-0.0974	0.0076	-0.0709	0.0008
		Pocock	0.0628	0.0076	0.0628	0.0029
	2	OBF	-0.0647	0.0035	-0.0580	-0.0029
		Pocock	0.0628	-0.0078	0.0628	-0.0133
	3	OBF	-0.0605	-0.0003	-0.0649	-0.0071
		Pocock	0.0628	-0.0106	0.0628	-0.0122
	4	OBF	-0.0582	-0.0009	-0.0444	0.0012
		Pocock	0.0628	-0.0040	0.0628	-0.0058

Table 6.8: Comparison of 5-look 2-times increment random GS design with HP005 boundaries versus equal-spaced random sample size re-estiamtion designs with OBF and PK boundaries when treatment effect follows Beta(5,2)

Total Looks	Reestimation Look	Method	HP005-Weighted		HP005-Unweighted	
			Difference of R_f	Difference of ABLC	difference of R_f	Difference of ABLC
2	1	OBF	-0.1284	-0.0102	-0.1269	-0.0042
		Pocock	-0.1087	-0.0101	-0.1080	-0.0122
5	1	OBF	-0.0659	-0.0064	-0.0394	-0.0132
		Pocock	0.0943	-0.0064	0.0943	-0.0111
	2	OBF	-0.0332	-0.0105	-0.0265	-0.0169
		Pocock	0.0943	-0.0218	0.0943	-0.0273
	3	OBF	-0.0290	-0.0143	-0.0334	-0.0211
		Pocock	0.0943	-0.0246	0.0943	-0.0262
	4	OBF	-0.0267	-0.0149	-0.0129	-0.0128
		Pocock	0.0943	-0.0180	0.0943	-0.0198

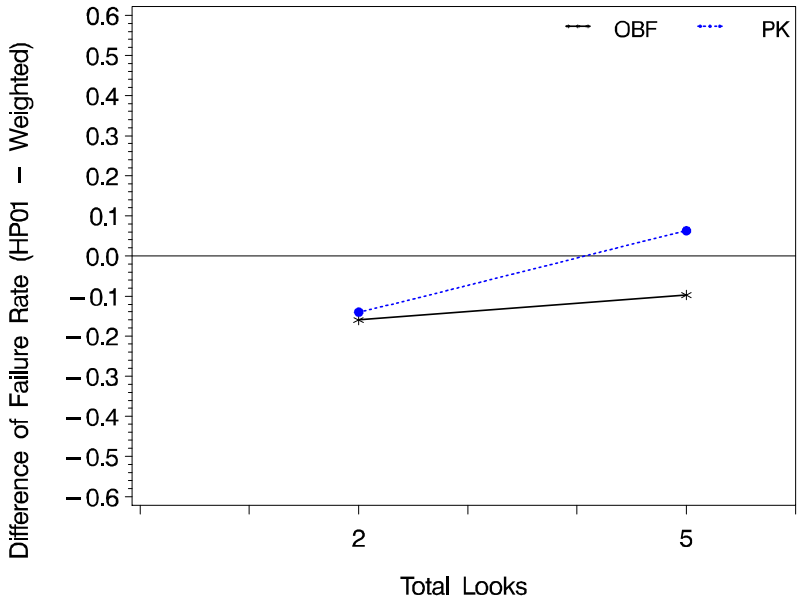

Figure 6.2: Failure Rate of 5-look Unequal-spaced GS Design using HP01 Boundary versus Weighted Designs using OBF & PK Boundary when Treatment Effect Follows Beta(5, 2)

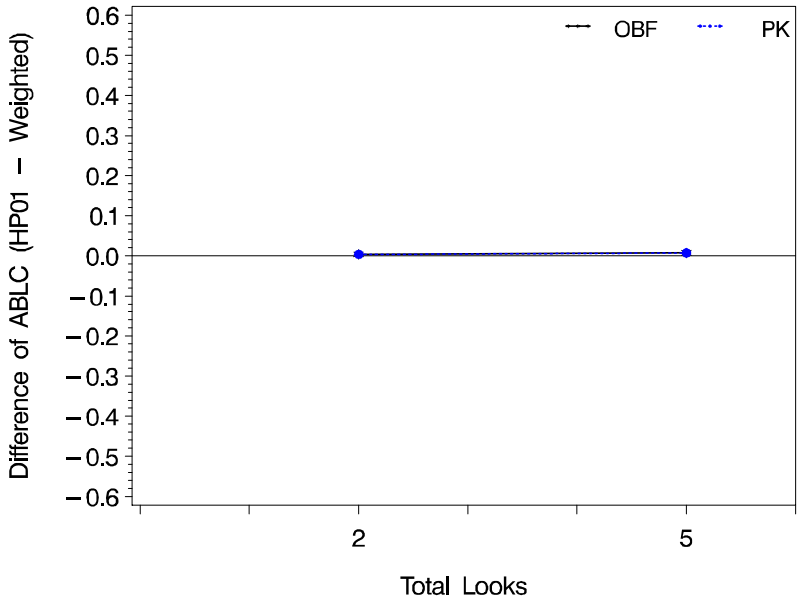

Figure 6.3: ABLC of 5-look Unequal-spaced GS Design using HP01 Boundary versus Weighted Designs using OBF & PK Boundary when Treatment Effect Follows Beta(5, 2)

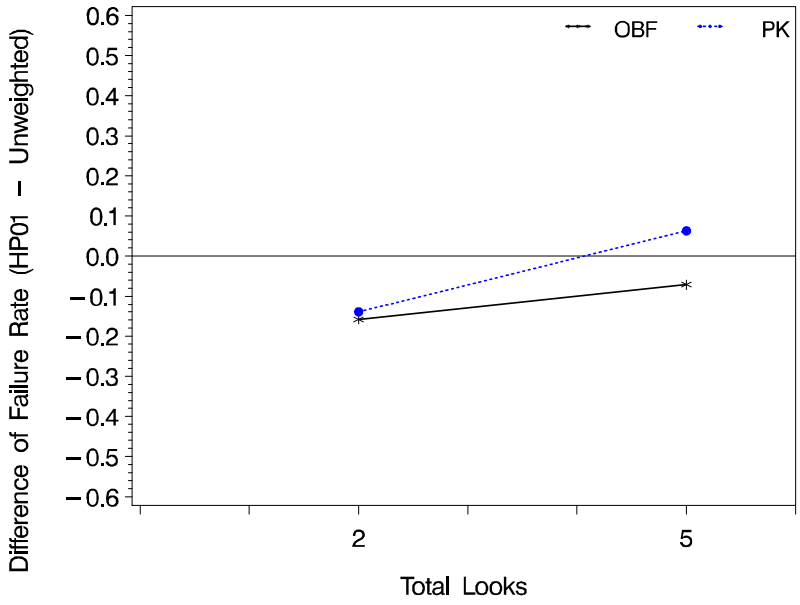

Figure 6.4: Failure Rate of 5-look Unequal-spaced GS Design using HP01 Boundary versus Unweighted Designs using OBF & PK Boundary when Treatment Effect Follows Beta(5, 2)

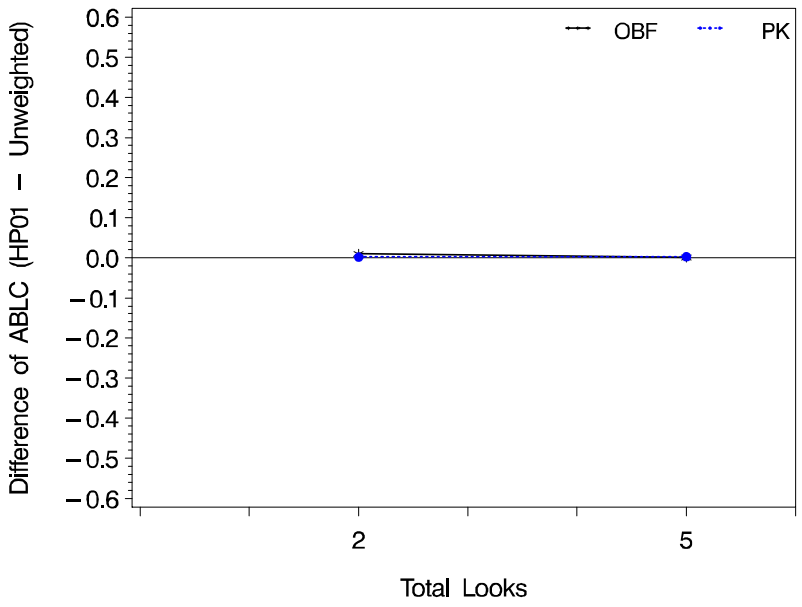

Figure 6.5: ABLC of 5-look Unequal-spaced GS Design using HP01 Boundary versus Unweighted Designs using OBF & PK Boundary when Treatment Effect Follows Beta(5, 2)

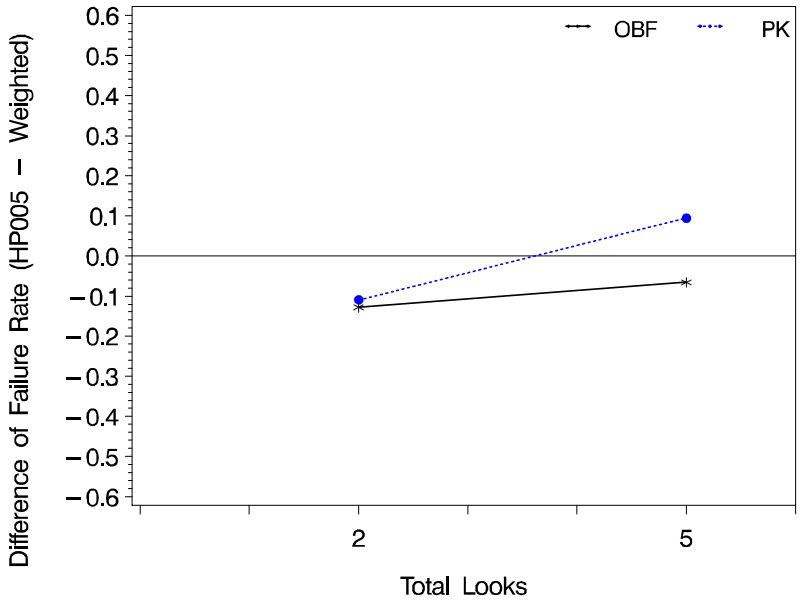

Figure 6.6: Failure Rate of 5-look Unequal-spaced GS Design using HP005 Boundary versus Weighted Designs using OBF & PK Boundary when Treatment Effect Follows Beta(5, 2)

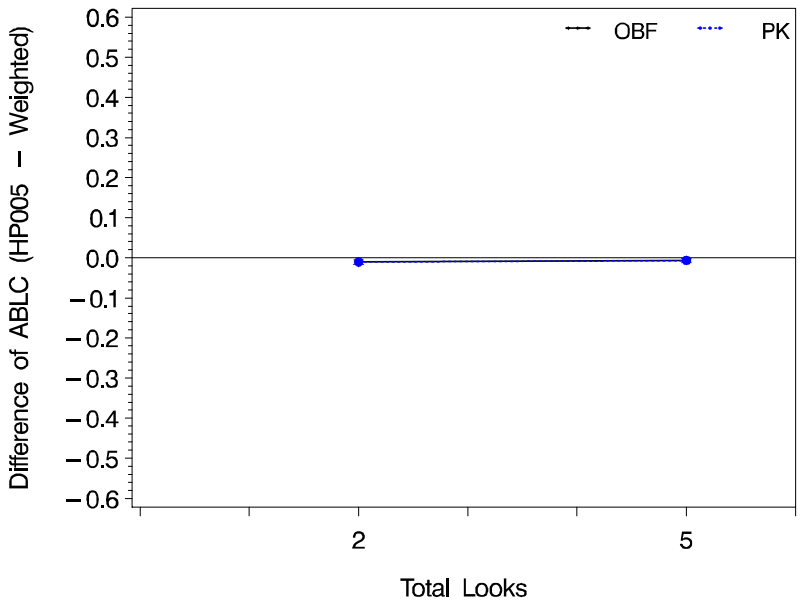

Figure 6.7: ABLC of 5-look Unequal-spaced GS Design using HP005 Boundary versus Weighted Designs using OBF & PK Boundary when Treatment Effect Follows Beta(5, 2)

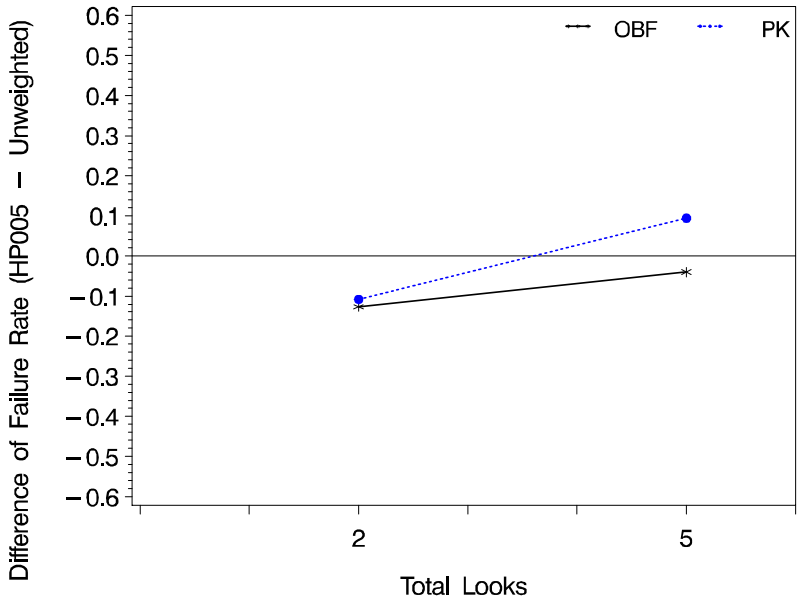

Figure 6.8: Failure Rate of 5-look Unequal-spaced GS Design using HP005 Boundary versus Unweighted Designs using OBF & PK Boundary when Treatment Effect Follows Beta(5, 2)

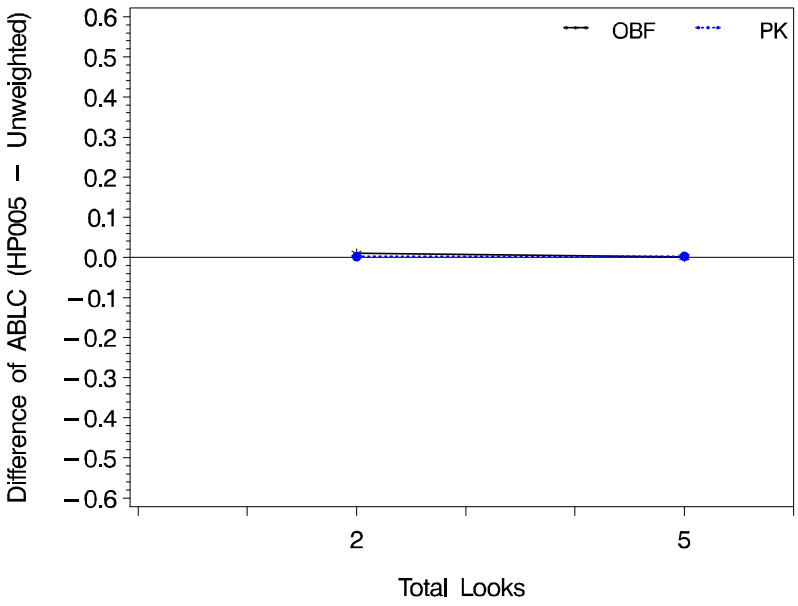

Figure 6.9: ABLC of 5-look Unequal-spaced GS Design using HP005 Boundary versus Unweighted Designs using OBF & PK Boundary when Treatment Effect Follows Beta(5, 2)

B) Treatment effect follows Beta(2,5)

Table 6.9: Comparison of 5-look 2-times increment random GS designs with HP01 boundaries versus equal-spaced random sample size re-estiamtion designs with OBF and PK boundaries when treatment effect follows Beta(2,5)

Total Looks	Reestimation Look	Method	HP01-Weighted		HP01-Unweighted	
			Difference of R_f	Difference of ABLC	difference of R_f	Difference of ABLC
2	1	OBF	-0.1232	-0.0809	-0.1240	-0.0840
		Pocock	-0.1266	-0.0693	-0.1124	-0.0618
5	1	OBF	-0.0312	-0.0319	0.0465	-0.0155
		Pocock	0.2437	0.0396	0.1989	0.0353
	2	OBF	0.0841	0.0073	0.1351	0.0067
		Pocock	0.3344	0.0617	0.2706	0.0572
	3	OBF	0.0929	0.0134	0.1420	0.0125
		Pocock	0.4551	0.0757	0.4110	0.0648
	4	OBF	0.0531	0.0039	0.1191	0.0123
		Pocock	0.2416	0.0329	0.2425	0.0347

Table 6.10: Comparison of 5-look 2-times increment random GS designs with HP005 boundaries versus equal-spaced random sample size re-estiamtion designs with OBF and PK boundaries when treatment effect follows Beta(2,5)

Total Looks	Reestimation Look	Method	HP005-Weighted		HP005-Unweighted	
			Difference of R_f	Difference of ABLC	difference of R_f	Difference of ABLC
2	1	OBF	-0.0102	-0.0365	-0.0110	-0.0396
		Pocock	-0.0136	-0.0249	0.0006	-0.0174
5	1	OBF	0.0818	0.0125	0.1595	0.0289
		Pocock	0.3567	0.0840	0.3119	0.0797
	2	OBF	0.1971	0.0517	0.2481	0.0511
		Pocock	0.4474	0.1061	0.3836	0.1016
	3	OBF	0.2059	0.0578	0.2550	0.0569
		Pocock	0.5681	0.1201	0.5240	0.1092
	4	OBF	0.1661	0.0483	0.2321	0.0567
		Pocock	0.3546	0.0773	0.3555	0.0791

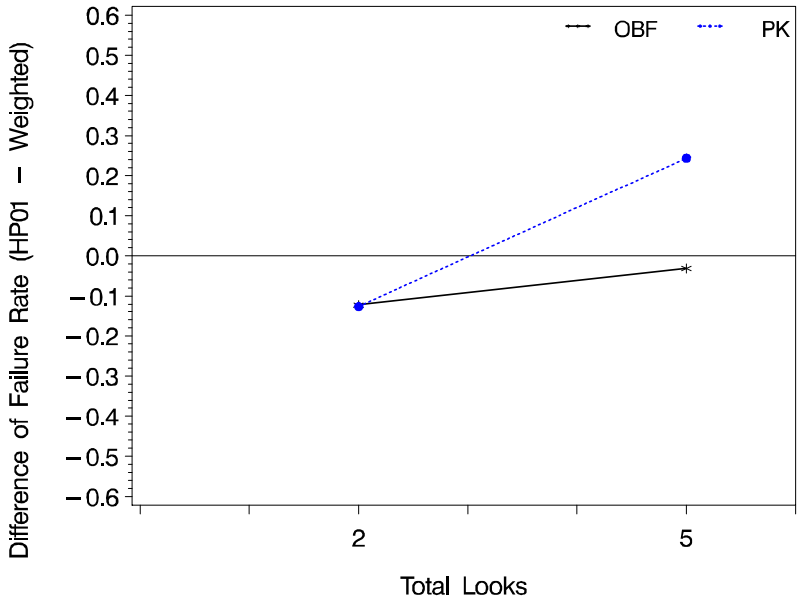

Figure 6.10: Failure Rate of 5-look Unequal-spaced GS Design using HP01 Boundary vs Weighted Designs using OBF & PK Boundary when Treatment Effect Follows Beta(2,5)

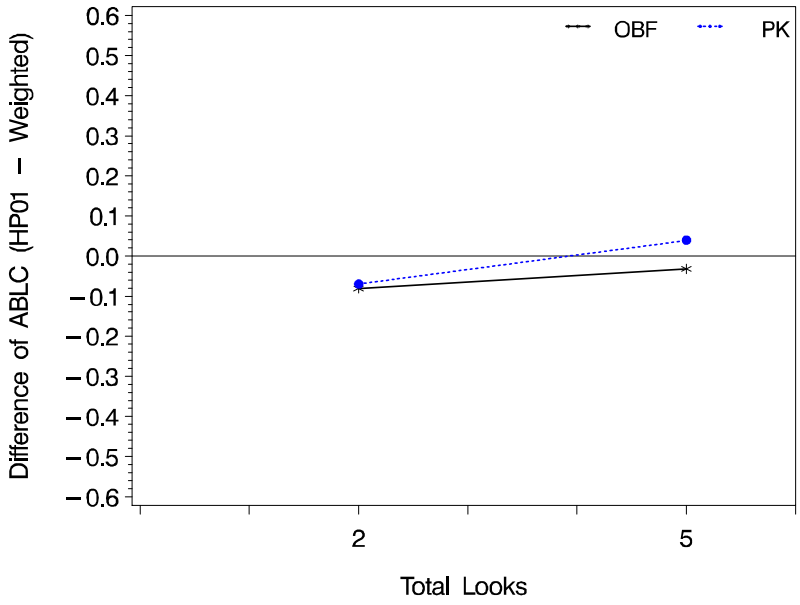

Figure 6.11: ABLC of 5-look Unequal-spaced GS Design using HP01 Boundary versus Weighted Designs using OBF & PK Boundary when Treatment Effect Follows Beta(2,5)

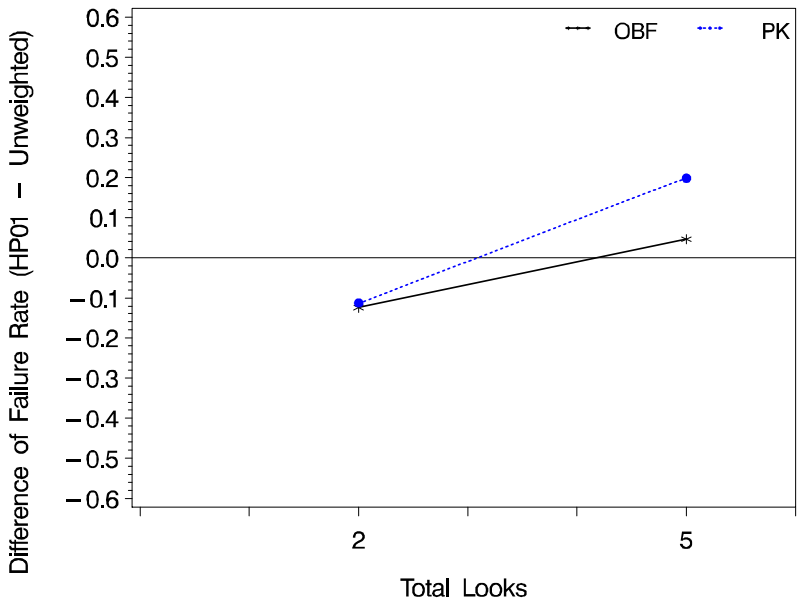

Figure 6.12: Failure Rate of 5-look Unequal-spaced GS Design using HP01 Boundary versus Unweighted Designs using OBF & PK Boundary when Treatment Effect Follows Beta(2,5)

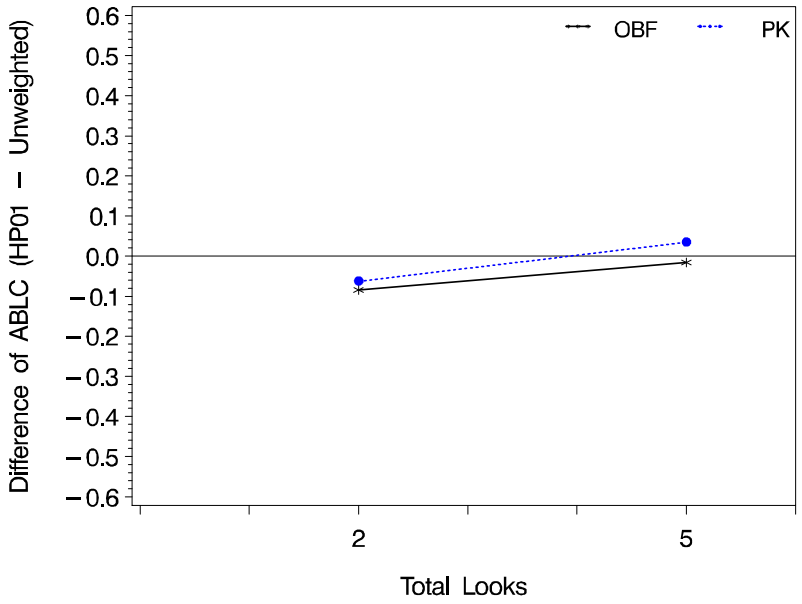

Figure 6.13: ABLC of 5-look Unequal-spaced GS Design using HP01 Boundary versus Unweighted Designs using OBF & PK Boundary when Treatment Effect Follows Beta(2,5)

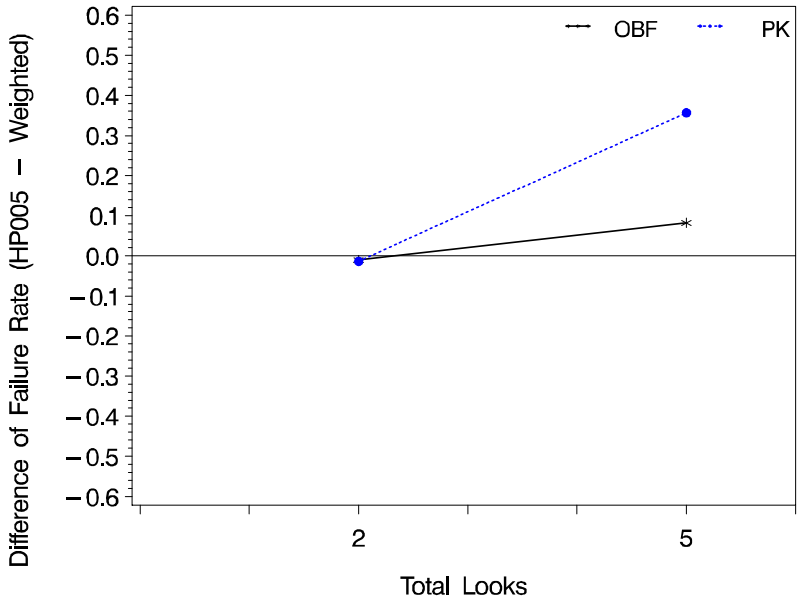

Figure 6.14: Failure Rate of 5-look Unequal-spaced GS Design using HP005 Boundary versus Weighted Designs using OBF & PK Boundary when Treatment Effect Follows Beta(2,5)

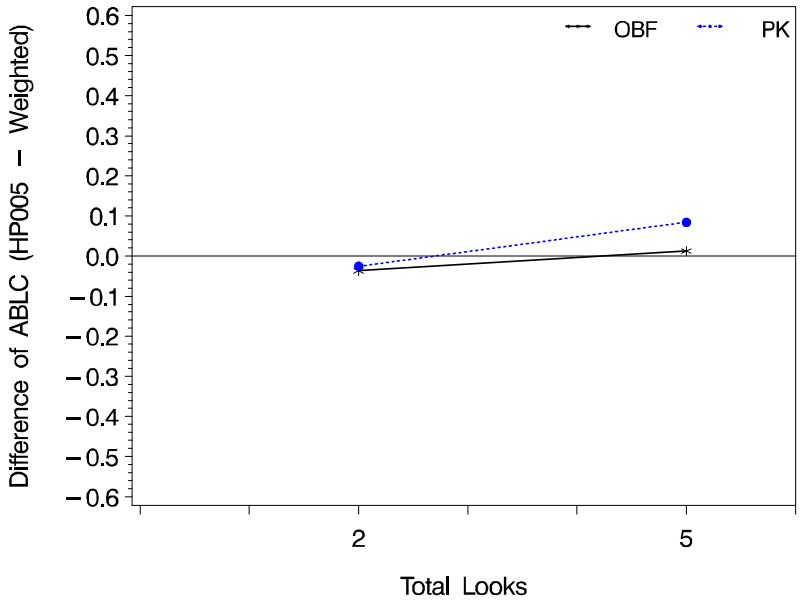

Figure 6.15: ABLC of 5-look Unequal-spaced GS Design using HP005 Boundary versus Weighted Designs using OBF & PK Boundary when Treatment Effect Follows Beta(2,5)

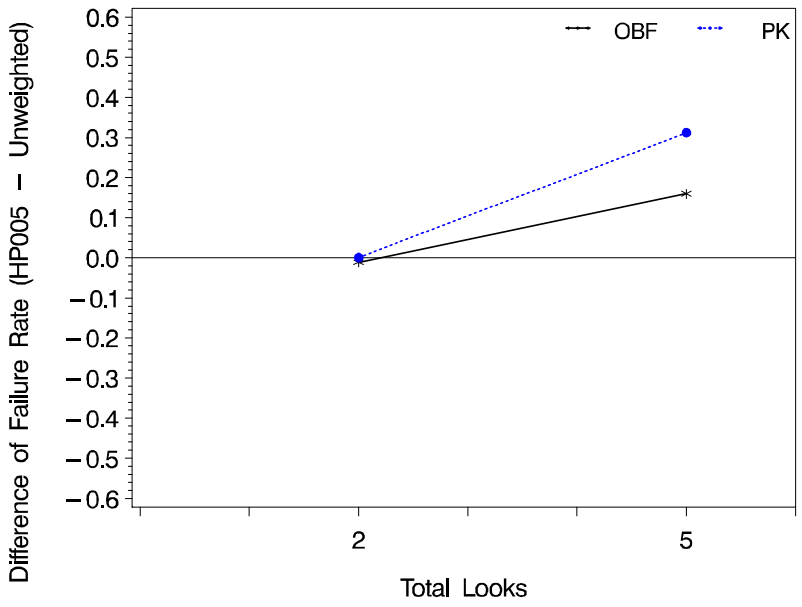

Figure 6.16: Failure Rate of 5-look Unequal-spaced GS Design using HP005 Boundary versus Unweighted Designs using OBF & PK Boundary when Treatment Effect Follows Beta(2,5)

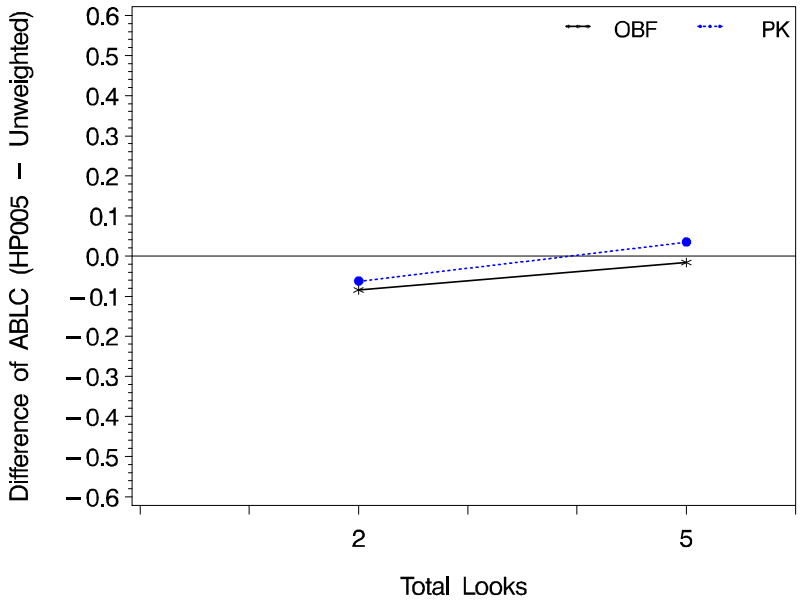

Figure 6.17: ABLC of 5-look Unequal-spaced GS Design using HP005 Boundary versus Unweighted Designs using OBF & PK Boundary when Treatment Effect Follows Beta(2,5)

C) Treatment effect follows Beta(4,5)

Table 6.11: Comparison of 5-look 2-times increment random GS designs with HP01 boundaries versus equal-spaced random sample size re-estiamtion designs with OBF and PK boundaries when treatment effect follows Beta(4,5)

Total Looks	Reestimation Look	Method	HP01-Weighted		HP01-Unweighted	
			Difference of R_f	Difference of ABLC	difference of R_f	Difference of ABLC
2	1	OBF	-0.1754	-0.1067	-0.1750	-0.0936
		Pocock	-0.1956	-0.1028	-0.1897	-0.0902
5	1	OBF	-0.0282	-0.0297	-0.0044	-0.0143
		Pocock	0.2236	0.0314	0.2774	0.0546
	2	OBF	0.0144	-0.0109	0.0340	0.0046
		Pocock	0.2124	0.0445	0.2880	0.0543
	3	OBF	0.0066	-0.0058	0.0427	0.0045
		Pocock	0.1505	0.0336	0.2492	0.0466
	4	OBF	-0.1106	-0.0282	-0.1379	-0.0315
		Pocock	0.0239	-0.0111	0.0076	-0.0094

Table 6.12: Comparison of 5-look 2-times increment random GS designs with hp005 boundaries versus equal-spaced random sample size re-estiamtion designs with OBF and PK boundaries when treatment effect follows Beta(4,5)

Total Looks	Reestimation Look	Method	HP005-Weighted		HP005-Unweighted	
			Difference of R_f	Difference of ABLC	difference of R_f	Difference of ABLC
2	1	OBF	-0.0088	-0.0562	-0.0084	-0.0431
		Pocock	-0.0290	-0.0523	-0.0231	-0.0397
5	1	OBF	0.1384	0.0208	0.1622	0.0362
		Pocock	0.3902	0.0819	0.4440	0.1051
	2	OBF	0.1810	0.0396	0.2006	0.0551
		Pocock	0.3790	0.0950	0.4546	0.1048
	3	OBF	0.1732	0.0447	0.2093	0.0550
		Pocock	0.3171	0.0841	0.4158	0.0971
	4	OBF	0.0560	0.0223	0.0287	0.0190
		Pocock	0.1905	0.0394	0.1742	0.0411

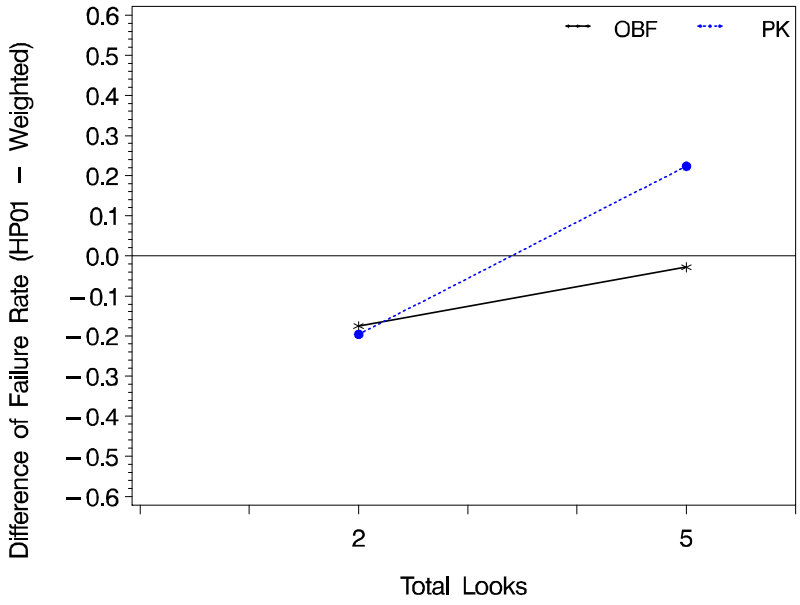

Figure 6.18: Failure Rate of 5-look Unequal-spaced GS Design using HP01 Boundary versus Weighted Designs using OBF & PK Boundary when Treatment Effect Follows Beta(4,5)

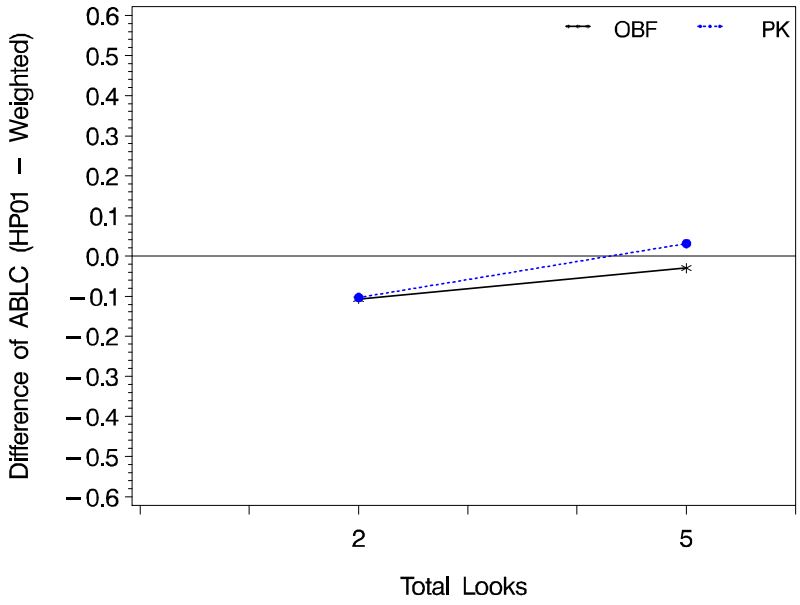

Figure 6.19: ABLC of 5-look Unequal-spaced GS Design using HP01 Boundary versus Weighted Designs using OBF & PK Boundary when Treatment Effect Follows Beta(4,5)

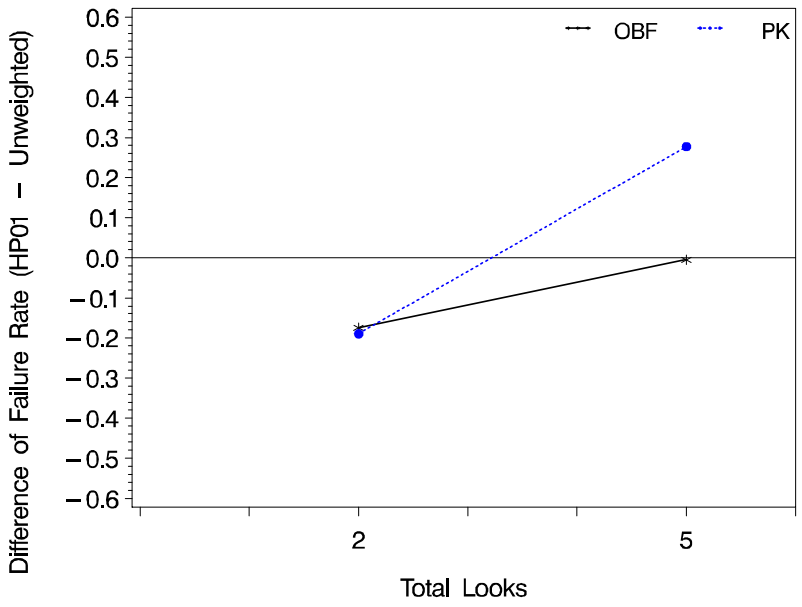

Figure 6.20: Failure Rate of 5-look Unequal-spaced GS Design using HP01 Boundary versus Unweighted Designs using OBF & PK Boundary when Treatment Effect Follows Beta(4,5)

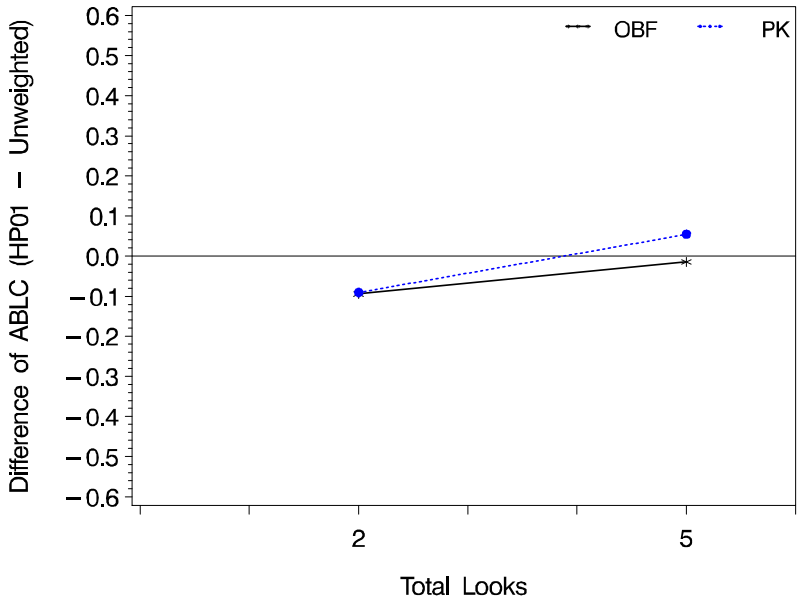

Figure 6.21: ABLC of 5-look Unequal-spaced GS Design using HP01 Boundary versus Unweighted Designs using OBF & PK Boundary when Treatment Effect Follows Beta(4,5)

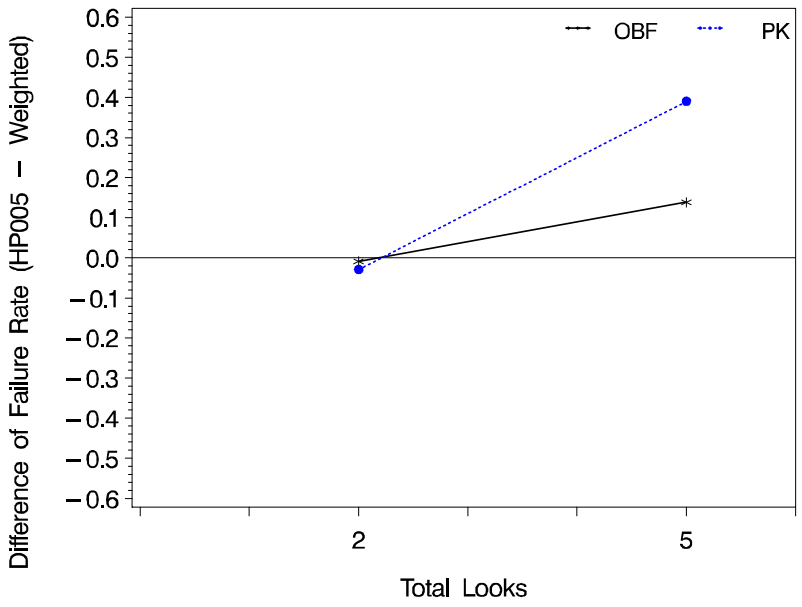

Figure 6.22: Failure Rate of 5-look Unequal-spaced GS Design using HP005 Boundary versus Weighted Designs using OBF & PK Boundary when Treatment Effect Follows Effect Beta(4,5)

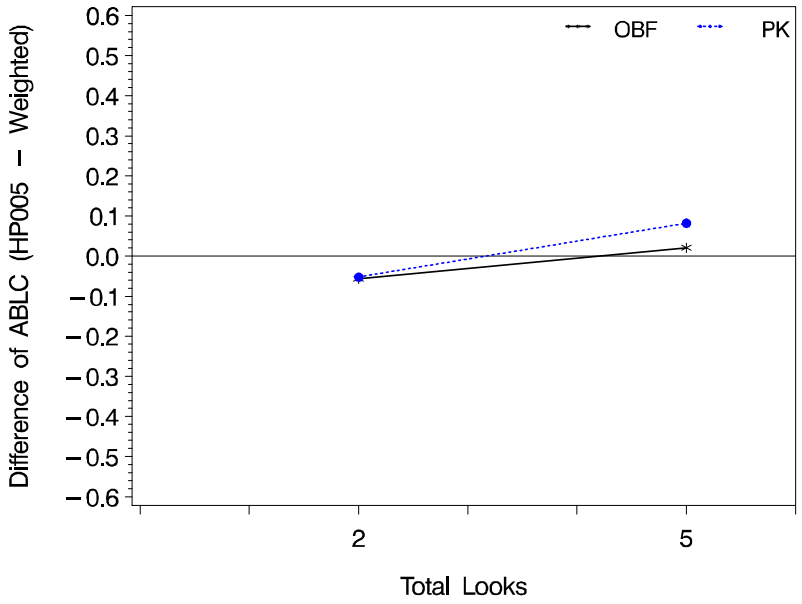

Figure 6.23: ABLC of 5-look Unequal-spaced GS Design using HP005 Boundary versus Weighted Designs using OBF & PK Boundary when Treatment Effect Follows Beta(4,5)

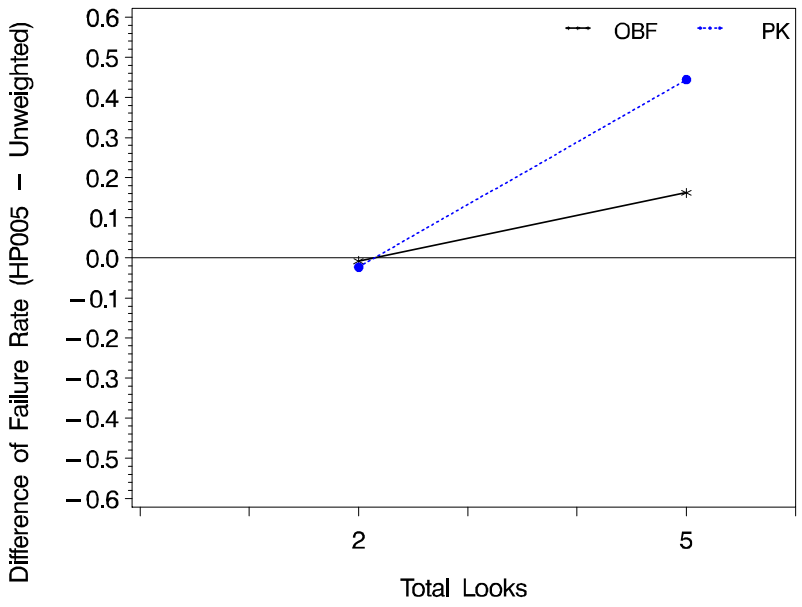

Figure 6.24: Failure Rate of 5-look Unequal-spaced GS Design using HP005 Boundary versus Unweighted Designs using OBF & PK Boundary when Treatment Effect Follows Effect Beta(4,5)

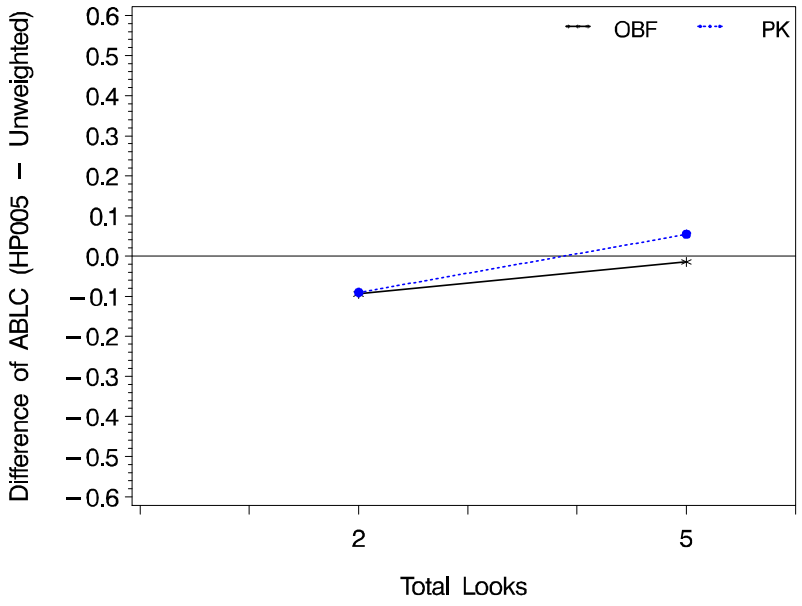

Figure 6.25: ABLC of 5-look Unequal-spaced GS Design using HP005 Boundary versus Unweighted Designs using OBF & PK Boundary when Treatment Effect Follows Beta(4,5)

Chapter 7

Conclusion

In the part I of this dissertation, the expected sample size for two-stage sample size re-estimation designs were derived, the golden standard for performance comparison was identified, the failure rate and the area between log curves as the performance measurements were defined, and the performance of adaptive designs when treatment effect follows a uniform or a beta distribution were compared.

The formula for expected sample size for two-stage weighted sample size re-estimation designs were derived in part I of this dissertation. Different formulas were provided for designs with or without the restrictions on the maximum sample size. Theoretical calculations through numerical method and simulations were done for both the sample size re-estimation designs and the group sequential designs when the maximum allowed sample size was 2018. Simulation results were very close to the results from the theoretical calculation. The expected sample size for unweighted sample size re-estimation designs was also discussed in part I. However, because more assumptions are required, to

derive expected sample size for unweighted sample size re-estimation designs will be considered as a future work.

Measurements of performance were also proposed in part I of this dissertation. Based on decision theory, failure rate and area between the log of sample size or power curves of adaptive designs and true treatment effect function were defined to evaluate the performance of adaptive designs on a treatment effect interval. Simulations were done to obtain the average sample size and power. Performance was measured based on different design parameters including different treatment effect intervals, maximum sample size, initial sample size for re-estimation designs, total number of looks, types of information increment, time of sample size re-estimation, etc. Treatment effect was assumed to follow either a uniform distribution or a beta distribution.

When the two-stage group sequential designs and sample size re-estimation designs with initial sample size of 356 were compared, both the failure rate and the ABLC were much smaller in sample size re-estimation designs. This is because the interim analysis for GS design was at half of the maximum sample size which was 1009. For those large treatment effects on the interval, the real sample sizes needed as shown on the true treatment effect function curve was much smaller than 1009. Comparisons of GS designs with different increments indicate that when other design parameters were the same, the performance for designs with 2 time unequal-spaced increment was better than the performance of GS designs with equal-spaced increment. The more the total looks was, the more the performance improvement in unequal spaced designs. For the comparison of two-stage group sequential designs (i.e. sample size re-estimation designs with re-estimation done

at the first interim look), the performance for weighted designs was almost identical to the performance of unweighted designs.

The most interesting comparison done was to compare the 5-look unequal-spaced GS design with HP type boundaries with sample size re-estimation designs with OBF and PK boundaries. It is known that the use of group sequential designs in clinical trials is well established. However, for sample size re-estimation designs, in a most recently published FDA guideline for adaptive designs, it is clearly indicated that adaptation of sample size based in the interim treatment effect estimates is still a less understood area [13]. So, by comparing performance of GS designs with different design parameters, if a GS design that can achieve the same as or better performance than sample size re-estimation designs can be found, it will be very meaningful. Results show that 5-looks unequal-spaced GS designs with HP01 and HP005 boundaries can achieve similar or better performance than the sample size re-estimation designs. Especially when the total number of looks for re-estimation designs is small, performance improvement is larger. However, since more interim analysis need to be done, it will require more careful planning on the timing and resources in advance when group sequential design with more interim looks is used.

Adaptive designs may not always be the best choice. Exploratory analysis indicates that when the treatment effect interval was very narrow, i.e. when there is a relatively accurate estimation of treatment effect, performance will be very robust on the interval regardless of the design chosen. Thus, a fixed sample size design may be good for some circumstances. However, because of the difficulty of obtaining such a narrow treatment effect interval, one should be cautious and may use simulations to confirm the point esti-

mate before the fixed sample size design instead of adaptive design is used.

Part II

Methods of Designing Two-stage Winner Designs with Survival Outcomes

Chapter 8

Introduction

8.1 Background

A common practice in drug development is to conduct a phase 2 study to provide dose and frequency information for a large scale phase 3 confirmatory trials. However, dependent on the nature of the disease, a phase 2 trial may not be feasible due to the long follow-up or the lack of resources. So it may be necessary to evaluate several promising regimens in the confirmatory phase 3 trial. In this case, an interim analysis is often used to drop the inferior arms and to avoid high cost, long term trial conduction, and patient exposure to ineffective treatments. This approach is considered as combining the two phases into one study: phase 2 portion will be carried out by the interim analysis. When appropriate surrogate endpoints exist, such as progression free survival in oncology trials, they can be used at the interim analysis to accelerate the drug development.

The study design of interest in this research was a two-stage winner design. In such a design (Figure 8.1), the study starts with two treatment groups and a control group.

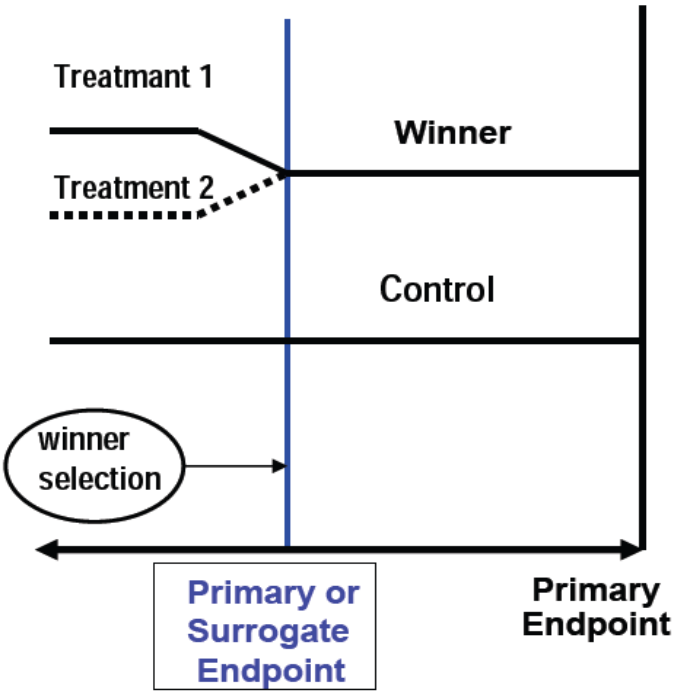

Figure 8.1: Two-stage Winner Design

The two treatment groups are compared at the interim analysis and only one treatment arm is allowed to enter the second stage of study. At the second stage, accrual continues onto the winning group and the control group till possibly the end of the study. The final analysis was based on all data in winner and control arms. Endpoint used at the first stage could be the same as or different from the primary endpoint used for final analysis. Or surrogate endpoints can be used in the first stage.

8.2 Research Objectives

Objective I: To derive the asymptotic correlation between log-rank statistics when the same endpoint is used at the interim and final analysis.

Objective II: To derive the asymptotic correlation between log-rank statistics when different endpoints are used at the interim and final analysis.

Objective III: To identify the formulas for approximate design parameter calculation and evaluate the accuracies of the approximations through simulations for two-stage winner designs when the same endpoint is used at interim and final analysis.

Objective IV: To identify the formulas for approximate design parameter calculation and evaluate the accuracies of the approximations through simulations for two-stage winner designs when different endpoints are used at interim and final analysis.

Chapter 9

Literature Review

With increasing interest in winner selection strategy, several designs were proposed. Shun et al [35] proposed a two-stage winner design with two treatments and a control group at the beginning. The inferior arm was dropped at the interim and only the winner arm and the control group carried over to the second stage. Continuous endpoints were used in this design. Exact method, as well as normal approximation, is provided for the tail probability, power, and sample size calculations. Since log-rank test statistics for survival outcomes were approximately normally distributed, the statistical frameworks for this design can also be used for two-stage designs with survival outcomes. Dunnett-type [7] testing procedure was proposed to compare the survival distribution of each experimental arm with the control arm. Jung et al. [17] proposed a procedure to choose a common critical value to control the family-wise error probability and a sample size calculation method using the log-rank statistics. Schaid et al., [33] offered an efficient two-stage design which screens out those new regimens not demonstrating a minimum pre-specified

survival advantage over standard regimen. Royston et al., [32] presented a design utilizing a surrogate to compare the experimental arms with the control at the first stage and comparing the winner arm with the control on the outcome measure of primary interest at the second stage. Progression free survival (PFS) was used for interim analysis and overall survival (OS) used for final analysis.

9.1 Two-stage winner design for continuous endpoints

Two-stage winner design was first proposed by Lan et al [19, 35]. Such a design starts with one control and two treatment groups. At the end of stage one, interim analysis was done to select a better treatment to continue to the second stage with the control group. Either a surrogate or the primary endpoint was used in the interim analysis for the two-stage winner design are studies in [19, 35]. The critical assumption of this design was that the interim and final test statistics were normally or asymptotically normally distributed. A normal approximation approach was described to simplify the calculations. The results of those paper can be generalized to time-to-event data and binary data, as long as the test statistics were asymptotically normal. But, the estimation of the covariance between the test statistics is challenging for survival outcomes.

9.1.1 Distribution of test statistics

When continuous variables are used as surrogate endpoint at the interim analysis, denote the continuous measurements $\{X_i^{(j)}|i = 1, ..., n_1\}$ i.i.d. with a normal distribution $N\left(v_j^X, \sigma_X\right)$, $j = 0, 1, 2$ for the interim analysis and $\{Y_i^{(j)}|i = 1, ..., n\}$ i.i.d. with a normal

distribution $N\left(\mu_j^Y, \sigma_Y\right), j = 0, 1, 2$ for the final analysis, where $j = 0, 1, 2$ is used to indicate the control and the two treatment groups respectively. Assume the correlation between $\{X_i^{(j)} | i = 1, ..., n_1\}$ and $\{Y_i^{(j)} | i = 1, ..., n\}$ is ρ and the information time is $\tau = n_1/n$, where n_1 and n are the sample size at interim and final analysis. The estimations for v_j^X and μ_j^Y are $\overline{X}_{n_1}^{(j)} = (1/n_1)\sum_{i=1}^{n_1} X_i^{(j)}$ and $\overline{Y}_n^{(j)} = (1/n)\sum_{i=1}^{n} X_i^{(j)}$, $for\ j = 0, 1, 2$.

At the interim, winner selection will be done based on the numeric value of $\overline{X}_{n_1}^{(1)}$ and $\overline{X}_{n_1}^{(2)}$. When $\overline{X}_{n_1}^{(1)} > \overline{X}_{n_1}^{(2)}$, treatment 1 will be selected. Otherwise treatment 2 will be selected. The hypotheses are $H_0 : \Delta_1 = \Delta_2 = 0$ versus $H_a : \Delta_1 > 0$ or $\Delta_2 > 0$, where $\Delta_j = \mu_j^Y - \mu_0^Y$ is the treatment different between j^{th} treatment and the control group. Let δ_j be the estimation of Δ_j. Let the test statistics

$$Z_n^{(j)} = \sqrt{\frac{n}{2\sigma_Y^2}}\left(\overline{Y}_n^{(j)} - \overline{Y}_n^{(0)}\right) = \sqrt{\tau} Z_{0,n_1}^{(j)} + \sqrt{1-\tau} Z_{n_1,n_2}^{(j)},$$

where $j = 1$ or 2, $Z_{0,n_1}^{(j)} = \frac{1}{n_1}\left(\sum_{i=1}^{n_1} Y_i^{(1)} - \sum_{i=1}^{n_1} Y_i^{(0)}\right)$ is the test statistic based on the first n_1 patients, and $Z_{n_1,n_2}^{(j)} = \frac{1}{n-n_1}\left(\sum_{i=n_1+1}^{n} Y_i^{(1)} - \sum_{i=n_1+1}^{n} Y_i^{(0)}\right)$ is the test statistic based on the rest of the patients.

The test statistic at the interim is

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

and the final test statistic is

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

The covariance η between $Z_n^{(j)}$ and V_{n_1} can be derived as follow:

$$
\begin{aligned}
\eta &= cov\left(Z_n^{(1)}, V_{n_1}\right) \\
&= cov\left(\sqrt{\tau}Z_{n_1}^{(1)} + \sqrt{1-\tau}Z_l^{(1)}, V_{n_1}\right) \\
&= cov\left(\sqrt{\tau}Z_{n_1}^{(1)}, V_{n_1}\right) \\
&= cov\left(\sqrt{\frac{n_1}{2\sigma_Y^2}}\sqrt{\tau}\left(\frac{\sum_{i=1}^{n_1} Y_i^{(1)}}{n_1} - \frac{\sum_{i=1}^{n_1} Y_i^{(0)}}{n_1}\right), \sqrt{\frac{n_1}{2\sigma_X^2}}\left(\frac{\sum_{i=1}^{n_1} X_i^{(1)}}{n_1} - \frac{\sum_{i=1}^{n_1} X_i^{(2)}}{n_1}\right)\right) \\
&= cov\left(\sqrt{\frac{n_1}{2\sigma_Y^2}}\sqrt{\tau}\frac{\sum_{i=1}^{n_1} Y_i^{(1)}}{n_1}, \sqrt{\frac{n_1}{2\sigma_X^2}}\left(\frac{\sum_{i=1}^{n_1} X_i^{(1)}}{n_1} - \frac{\sum_{i=1}^{n_1} X_i^{(2)}}{n_1}\right)\right) \\
&= \sqrt{\frac{n_1}{2\sigma_Y^2}}\sqrt{\tau}\sqrt{\frac{n_1}{2\sigma_X^2}}\frac{1}{n_1^2}cov\left(\sum_{i=1}^{n_1} Y_i^{(1)}, \sum_{i=1}^{n_1} X_i^{(1)}\right) \\
&= \sqrt{\frac{n_1}{2\sigma_Y^2}}\sqrt{\tau}\sqrt{\frac{n_1}{2\sigma_X^2}}\frac{1}{n_1^2}n_1\rho\sigma_X\sigma_Y \\
&= \frac{\sqrt{\tau}}{2}\rho = -cov\left(Z_n^{(2)}, V_{n_1}\right)
\end{aligned}
$$

The distribution of W is

$$F_W(w) = pF_1(w - w_1) + qF_2(w - w_2),$$

and the density function of W is

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

where p is the winning probability, $w_1 = \sqrt{\frac{n}{2\sigma_Y^2}}\delta_1$, $w_2 = \sqrt{\frac{n}{2\sigma_Y^2}}\delta_2$. Since under $H_a : V_{n_1} \sim N(\lambda, 1)$ and $\lambda = \sqrt{\frac{n_1}{2\sigma_X^2}}\left(v_1^X - v_2^X\right)$, it can be calculated as $p = \Pr(V_{n_1} > 0) = \Pr(V_{n_1} - \lambda > -\lambda) = 1 - \Phi(-\lambda) = \Phi(\lambda)$. Also we have $q = 1 - p$, $f_1(w) = \frac{1}{p}\Phi(k_0 + kw)\phi(w)$, and $f_2(w) = \frac{1}{q}\Phi(-k_0 + kw)\phi(w)$. Let $W_1 \sim f_1$ and $W_2 \sim f_2$ be two random variables defined by f_1 and f_2. The means and variances of W_1 and W_2 under H_a are

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

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

respectively, where

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

Type I error rate can be computed as:

$$\Pr\left(W > Z_\alpha\right) = \Pr\left(Z_n^{(1)} > Z_\alpha, V_{n_1} > 0\right) + \Pr\left(Z_n^{(2)} > Z_\alpha, V_{n_1} < 0\right).$$

This probability can be computed numerically since under H_0 :

$$\begin{pmatrix} Z_n^{(1)} \\ V_{n_1} \end{pmatrix} \sim \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} 1 & \frac{\sqrt{\tau}}{2}\rho \\ \frac{\sqrt{\tau}}{2}\rho & 1 \end{bmatrix} \right)$$

and

$$\begin{pmatrix} Z_n^{(2)} \\ V_{n_1} \end{pmatrix} \sim \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} 1 & -\frac{\sqrt{\tau}}{2}\rho \\ -\frac{\sqrt{\tau}}{2}\rho & 1 \end{bmatrix} \right).$$

9.1.2 Normal Approximation

To simplify the calculation, a normal approximation approach is proposed. W_1 and W_2 can be approximated by the following normal random variables:

$$Z_1 \sim N\left(\mu_1, \sigma_1^2\right) \text{ and } Z_2 \sim N\left(\mu_2, \sigma_2^2\right).$$

And $f_W(w)$ can be approximated as

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

Type I error rate, power and sample size calculation can also be done based on the normal approximation.

9.2 Asymptotic distribution of unstandardized log-rank test statistics

Based on the work of Tsiatis [37, 38] the asymptotic joint distribution of unstandardized log rank statistics is given by Schaid et al. [33]. Use similar notations as in Schaid et al. [33]. The nonnegative random variables Y_i, V_i and W_i denote the real time of entry during the accrual period $[0, t_a]$, the time from entry until failure, and the time to censor, respectively. Assume the total accrual is at a constant rate. Let the vector $\mathbf{Z}_i = (Z_{0i}, ..., Z_{Ki})$ indicate treatment assignment; $z_{ji} = 1$ if the i^{th} patient is assigned to treatment j, $z_{ji} = 0$ otherwise ($j = 0, 1, ..., K$; z_{0i} represents control). Denote the distribution of entry times and censoring times as $H(y) = P(Y_i \leq y)$ and $G(w) = P(W_i \leq w)$ respectively, and define $\bar{G}(w) = 1 - G(w)$. The hazard rate function for the time from entry until failure is $\lambda_0(t)$ for the control arm and the cumulative hazard function is $\Lambda_0(t) = \int_0^t \lambda_0(x)dx$. Furthermore, let

$$X_i(t) = \max\{\min(V_i, t - Y_i, W_i), 0\},$$

$$\triangle_i(t) = I_{[V_i < \min(t - Y_i, W_i)]},$$

where $I_{[A]}$ denotes the indicator function of event A.

The log-rank statistic, not standardized by its variance, for the comparison at time t of treatment j versus control may be written as

$$S_N^j(t) = \sum_{i=1}^{N} (z_{0i} + z_{ji}) \triangle_i(t) [z_{ji} - \bar{z}_j\{t, X_i(t)\}],$$

where N denotes the maximum total number of patients in the study,

$$\bar{z}_j(t,x) = \sum_{i \in R_k(t,x)} z_{ji} / \sum_{i \in R_k(t,x)} I_{[X_i(t) \geq x]},$$

and $R_j(t,x)$ denotes the risk set for patients in either treatment j or control with $X_i(t) \geq x$. it is shown that

$$N^{-1/2}\{S_N^1(t_1), S_N^1(t_2), ..., S_N^k(t_1), S_N^k(t_2)\}$$

has an asymptotic multivariate normal distribution [38]. Under $H_0, E\{N^{-1/2}S_N^j(t)\} = 0$, the asymptotic covariance between $N^{-1/2}S_N^j(t)$ and $N^{-1/2}\{S_N^{j'}(t')$ are

$$\sigma^{jj'}\{t,t'\} = \psi(t)(\nu_0\nu_j^2 + \nu_0^2\nu_j)/(\nu_0+\nu_j)^2, \text{when } j = j' \text{and } t \leq t'$$

$$\sigma^{jj'}\{t,t'\} = \psi(t)\nu_0\nu_j\nu_j'/(\nu_0+\nu_j)(\nu_0+\nu_j'), \text{when } j \neq j' \text{and } t \leq t'$$

where

$$\psi(t) = \int_0^t \lambda_0(x)\exp(-\Lambda_0(x)\}H(t-x)\bar{G}(x)dx, \quad \nu_j = E(z_j) \qquad (j = 0, 1, ..K).$$

In Tsiatis [37], for without censoring process, the unstandardized log-rank test statistic is denoted as $D\,(t)$. Let

$$\overline{D}\,(t) = \sum_{i=1}^{N}(Z_i - \mu_Z)\,[\triangle_i(t) - \Lambda\{X_i(t)\}]\,,$$

where $\Lambda(s)$ is the cumulative hazard function defined as

$$\Lambda(s) = \begin{cases} \int_0^s \lambda(y)dy & \text{if } s \geq 0, \\ 0 & \text{if } s < 0. \end{cases}$$

The asymptotic joint distribution of $N^{1/2}\left\{D\left(t\right),D\left(t'\right)\right\}$ is the same as the asymptotic joint distribution of $N^{1/2}\left\{\overline{D}\left(t\right),\overline{D}\left(t'\right)\right\}$, which is equal to

$$N^{1/2}\left[\sum_{i=1}^{N}(Z_i-\mu_Z)\left[\triangle_i(t)-\Lambda\{X_i(t)\}\right],\sum_{i=1}^{N}(Z_i-\mu_Z)\left[\triangle_i(t')-\Lambda\{X_i(t')\}\right]\right]. \tag{9.1}$$

Chapter 10

Asymptotic correlation of log-rank statistics in two-stage winner design

In drug development, especially the development of treatments for cancer, surrogate endpoints are often used as a substitute for the primary endpoint. Since the distribution of log-rank test statistics for survival endpoints are asymptotically normal, the critical assumption of two-stage winner design is still valid. The statistical framework for two-stage winner design using continuous endpoints can be used for survival endpoints. As the calculations in the normal approximation approach is based on a bivariate normal distribution, the most challenging part in two-stage winner design using survival endpoints is how to estimate the covariance between interim and final log-rank statistics. In the next sections, extended work on two-stage winner design using survival endpoints is presented.

10.1 Asymptotic correlation of log-rank test statistics

Based on the work of Tsiatis [37, 38], when there are more than two treatment groups in the study, the asymptotic joint distribution of unstandardized log rank statistics is given by Schaid et al [33]. However, the log rank statistics Schaid studied was based on the same endpoint at different time points. When the same survival endpoint was used at the interim and final analysis, the derivation of asymptotic correlation of log rank statistics at interim and final was straightforward. But, when surrogate endpoint was used at interim, there was no covariance matrix given in the literature. In order to estimate the type I error inflation in two-stage winner design, covariance matrix is derived below.

Use similar notations as in Schaid et al. [33]. Let the nonnegative random variables Y_i, V_i and W_i denote the real time of entry of i^{th} patient during the accrual period $[0, t_a]$, the time from entry until failure, and the time to censor, respectively. Assume there are two treatment groups in addition to the control group. Let the vector $\mathbf{Z}_i$ with three elements indicate treatment assignment; $z_{ji} = 1$ if the i^{th} patient is assigned to treatment j, $z_{ji} = 0$ otherwise ($j = 0, 1, 2$; z_{0i} represents control). And the allocation ratio is $\nu_0 : \nu_1 : \nu_2 = 1 : 1 : 1$, where $\nu_0 + \nu_1 + \nu_2 = 1$. Denote the distribution of entry times and censoring times as $H(y) = P(Y_i \leq y)$ and $G(w) = P(W_i \leq w)$ respectively, and define $\bar{G}(w) = 1 - G(w)$. The hazard rate function for the time from entry until failure is $\lambda_0(t)$ for the control arm and the cumulative hazard function is $\Lambda_0(t) = \int_0^t \lambda_0(x)dx$. Furthermore,

$$X_i(t) = \max\{\min(V_i, t - Y_i, W_i), 0\},$$

$$\triangle_i(t) = I_{[V_i < \min(t - Y_i, W_i)]},$$

where $I_{[A]}$ denotes the indicator function of event A.

The log-rank statistic, not standardized by its variance, for the comparison at time t of control versus treatment j may be written as

$$S_N^j(t) = \sum_{i=1}^{N}(z_{0i} + z_{ji})\triangle_i(t)[z_{0i} - \bar{z}_j\{t, X_i(t)\}],$$

where N denotes the maximum total number of patients in the study,

$$\bar{z}_j(t,x) = \sum_{i\in R(t,x)}(z_{0i} + z_{ji})z_{0i} / \sum_{i\in R(t,x)}(z_{0i} + z_{ji})I_{[X_i(t)\geq x]},$$

and $R(t,x)$ denotes the risk set for all patients in the study with $X_i(t) \geq x$. Treatment j is considered better than the control if the statistic is greater than 0.

Denote the unstandardized log-rank statistic for comparison of treatment 2 versus treatment 1 at time t as

$$S_N^*(t) = \sum_{i=1}^{N}(z_{1i} + z_{2i})\triangle_i(t)[z_{2i} - \bar{z}^*\{t, X_i(t)\}],$$

where

$$\bar{z}^*(t,x) = \sum_{i\in R(t,x)}(z_{1i} + z_{2i})z_{2i} / \sum_{i\in R(t,x)}(z_{1i} + z_{2i})I_{[X_i(t)\geq x]},$$

and $R(t,x)$ denotes the risk set for all patients in the study with $X_i(t) \geq x$. Note that treatment 1 is better if the statistic is greater than 0.

10.1.1 Using the same endpoint at interim and final analysis

Under H_0, $E\{N^{-1/2}S_N^j(t)\} = 0$ and $E\{N^{-1/2}S_N^*(t)\} = 0$, $j = 1,2$. Based on Schaid et al (1990), the asymptotic variances are

$$\sigma^2\{N^{-1/2}S_N^j(t)\} = \psi(t)(\nu_0\nu_j^2 + \nu_0^2\nu_j)/(\nu_0 + \nu_j)^2, j = 1,2,$$

$$\sigma^2\{N^{-1/2}S_N^*(t)\} = \psi(t)(\nu_1\nu_2^2 + \nu_1^2\nu_2)/(\nu_1+\nu_2)^2,$$

where

$$\psi(t) = \int_0^t \lambda_0(x)\exp(-\Lambda_0(x)\}H(t-x)\bar{G}(x)dx, \quad \nu_j = E(z_j)\ (j=0,1,2).$$

The asymptotic covariances for $t \le t'$ are

$$cov\{N^{-1/2}S_N^1(t'), N^{-1/2}S_N^*(t)\} = \psi(t)\nu_1\nu_0\nu_2/\{(\nu_1+\nu_2)(\nu_1+\nu_0)\}$$

$$cov\{N^{-1/2}S_N^2(t'), N^{-1/2}S_N^*(t)\} = \psi(t)\nu_1\nu_0\nu_2/\{(\nu_2+\nu_1)(\nu_2+\nu_0)\}.$$

It can be derived from the above that the increment $N^{-1/2}(S_N^j(t') - S_N^j(t))$ is asymptotically independent of $N^{-1/2}S_N^*(t)$. From the variance and covariance formulas we have the asymptotic correlations

$$\begin{aligned}
&corr\{N^{-1/2}S_N^1(t'), N^{-1/2}S_N^*(t)\} \\
&= \frac{cov\{N^{-1/2}S_N^1(t'), N^{-1/2}S_N^*(t)\}}{\sigma\{N^{-1/2}S_N^1(t)\}\sigma\{N^{-1/2}S_N^*(t)\}} \\
&= \psi(t)\nu_1\nu_0\nu_2/\sqrt{\psi(t)(\nu_0\nu_1^2+\nu_0^2\nu_1)\psi(t')(\nu_1\nu_2^2+\nu_1^2\nu_2)} \\
&= \sqrt{\psi(t)/\psi(t')}\sqrt{\nu_0\nu_2/[(\nu_0+\nu_1)(\nu_1+\nu_2)]} \\
&corr\{N^{-1/2}S_N^2(t'), N^{-1/2}S_N^*(t)\} \\
&= \frac{cov\{N^{-1/2}S_N^2(t'), N^{-1/2}S_N^*(t)\}}{\sigma\{N^{-1/2}S_N^2(t)\}\sigma\{N^{-1/2}S_N^*(t)\}} \\
&= -\psi(t)\nu_1\nu_0\nu_2/\sqrt{\psi(t)(\nu_0\nu_2^2+\nu_0^2\nu_2)\psi(t')(\nu_1\nu_2^2+\nu_1^2\nu_2)} \\
&= -\sqrt{\psi(t)/\psi(t')}\sqrt{\nu_0\nu_1/[(\nu_0+\nu_2)(\nu_1+\nu_2)]}.
\end{aligned}$$

Similarly,

$$corr\{N^{-1/2}S_N^1(t), N^{-1/2}S_N^2(t)\}$$

$$= \frac{cov\{N^{-1/2}S_N^1(t), N^{-1/2}S_N^2(t)\}}{\sigma\{N^{-1/2}S_N^1(t)\}\sigma\{N^{-1/2}S_N^2(t)\}}$$
$$= \psi(t)\nu_1\nu_0\nu_2/\sqrt{\psi(t)(\nu_0\nu_1^2+\nu_0^2\nu_1)\psi(t)(\nu_0\nu_2^2+\nu_0^2\nu_2)}$$
$$= \sqrt{\nu_1\nu_2/[(\nu_0+\nu_1)(\nu_0+\nu_2)]}.$$

Let $LR^{(1')} = \frac{S_N^1(t)}{\sqrt{N\sigma^{11}}}$ and $LR^{(2')} = \frac{S_N^2(t)}{\sqrt{N\sigma^{22}}}$ denote the standardized log-rank test statistic at stage 1 and stage 2 respectively. Since

$$N^{-1/2}\left\{S_N^1(t), S_N^2(t)\right\} \xrightarrow{d} N\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}\sigma_{11} & \sigma_{12}\\ \sigma_{12} & \sigma_{22}\end{pmatrix}\right)$$

and

$$corr\left\{S_N^1(t), S_N^2(t)\right\} = \frac{cov\left\{N^{-1/2}S_N^1(t), N^{-1/2}S_N^2(t)\right\}}{\sqrt{var\left(N^{-1/2}S_N^1(t)\right)}\sqrt{var\left(N^{-1/2}S_N^2(t)\right)}}$$
$$= \frac{N\sigma^{12}}{N\sqrt{\sigma^{11}\sigma^{22}}} = \frac{\sigma^{12}}{\sqrt{\sigma^{11}\sigma^{22}}},$$

we can get the distribution for standardized log-rank statistics

$$\left\{LR^{(1')}, LR^{(2')}\right\} \xrightarrow{d} N\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & \frac{\sigma^{12}}{\sqrt{\sigma^{11}\sigma^{22}}}\\ \frac{\sigma^{12}}{\sqrt{\sigma^{11}\sigma^{22}}} & 1\end{pmatrix}\right).$$

By Slutsky's theorem , we know that $\left(LR^{(1)}, LR^{(2)}\right)$ have the same asymptotic distribution as $\left(LR^{(1')}, LR^{(2')}\right)$, where $LR^{(1)} = \frac{S_N^1(t)}{\sqrt{N\hat{\sigma}^{11}}}$ and $LR^{(2)} = \frac{S_N^2(t)}{\sqrt{N\hat{\sigma}^{22}}}$ for any consistent estimations of σ^{11} and σ^{22}. Thus the asymptotic correlation between the standardized statistics are $\frac{\sigma^{12}}{\sqrt{\sigma^{11}\sigma^{22}}}$, which is equal to the asymptotic correlation between the unstandardized log-rank statistics.

At the first stage, let $LR_{01}^{(1)}$ and $LR_{02}^{(1)}$ denote the log-rank statistics for treatment 1 versus control and treatment 2 versus control respectively. Similar to the notation used

in the Shun, Lan & Soo paper [35], use V to denote log-rank statistic for treatment 1 vs treatment 2.

Since if $t = t^{'}$, then $\sqrt{\psi(t)/\psi(t')} = 1$, the correlation

$$corr(LR_{01}^{(1)}, V) = \sqrt{\nu_0\nu_2/[(\nu_0+\nu_1)(\nu_1+\nu_2)]},$$

and

$$corr(LR_{02}^{(1)}, V) = -\sqrt{\nu_0\nu_1/[(\nu_0+\nu_2)(\nu_1+\nu_2)]}.$$

The correlation between $LR_{01}^{(1)}$ and $LR_{02}^{(1)}$ is

$$corr(LR_{01}^{(1)}, LR_{02}^{(1)}) = \sqrt{\nu_1\nu_2/[(\nu_0+\nu_1)(\nu_0+\nu_2)]}.$$

When the allocation ratio is 1:1:1, the correlations will be

$$corr(LR_{01}^{(1)}, V) = 1/2,$$

$$corr(LR_{02}^{(1)}, V) = -1/2,$$

and

$$corr(LR_{01}^{(1)}, LR_{02}^{(1)}) = 1/2.$$

At the second stage, we know the asymptotic correlation between $LR_{0j}^{(1)}$ and $LR_{0j}^{(2)}$ is $\sqrt{\tau}$, where $\tau = d_1/d_2$ is the information time. The relation between V and the increments of the second stage log-rank statistics I_j for treatment j versus control is

$$LR_{0j}^{(2)} = \sqrt{\tau}LR_{0j}^{(1)} + \sqrt{1-\tau}I_j,$$

and

$$I_j = [LR_{0j}^{(2)} - \sqrt{\tau} LR_{0j}^{(1)}] / \sqrt{1-\tau}, j = 1, 2.$$

We know

$$corr(I_1, V) = 0,$$

and

$$corr(I_2, V) = 0.$$

Then it can be shown, asymptotically, that

$$\begin{aligned} \eta_1 \equiv cov(LR_{01}^{(2)}, V) &= \sqrt{\tau} cov(LR_{01}^{(1)}, V) + \sqrt{1-\tau} cov(I_1, V) \\ &= \sqrt{\tau} cov(LR_{01}^{(1)}, V) \\ &= \sqrt{\tau \nu_0 \nu_2 / [(\nu_0 + \nu_1)(\nu_1 + \nu_2)]}, \end{aligned} \tag{10.1}$$

and

$$\begin{aligned} \eta_2 \equiv cov(LR_{02}^{(2)}, V) &= \sqrt{\tau} cov(LR_{02}^{(1)}, V) + \sqrt{1-\tau} cov(I_2, V) \\ &= \sqrt{\tau} cov(LR_{02}^{(1)}, V) \\ &= -\sqrt{\tau \nu_0 \nu_1 / [(\nu_0 + \nu_2)(\nu_1 + \nu_2)]}. \end{aligned} \tag{10.2}$$

When allocation ratio 1:1:1, we have

$$\eta_1 = \frac{\sqrt{\tau}}{2} = -\eta_2.$$

Now we know that $(V, LR_{01}^{(2)})$ and $(V, LR_{02}^{(2)})$ both approximately have bivariate normal distribution. They have unit variance and the covariances are given in Formulas 10.1 and 10.2. The critical value could be determined once τ is known. The power under

H_a will be

$$P(W > w) = P(V \geq 0, LR_{01}^{(2)} > w) + P(V < 0, LR_{02}^{(2)} > w).$$

With these results, one can design the trial using numerical bivariate normal cumulative distribution function or the normal approximation method of Shun et al [35].

10.1.2 Using surrogate endpoint at interim analysis

In clinical trial development, it may not be feasible to use primary endpoint to do the interim analysis due to the long follow up, such as survival. Thus, it is necessary to use a surrogate endpoint at the first stage. In this section, the correlation between log-rank statistics based on different survival endpoints is developed.

A. Covariance matrix for log rank statistics

Tsiatis [38] discussed the asymptotic distribution of test statistics in survival analysis. When there are two treatment groups (treatment versus control), the covariance matrix for the log-rank test statistics at two different time points ($t_1$and t_2) based on the same endpoint is

$$\Omega' = \begin{bmatrix} \sigma'_{11} & \sigma'_{12} \\ \sigma'_{12} & \sigma'_{22} \end{bmatrix},$$

where

$$\sigma'_{11} = \int_0^{t_1} E[\Big(z_j - \mu_{z_j}\Big)^2] H(t_1 - x)\bar{G}(x) \exp\{-\Lambda_1(x)\}\lambda_1(x)dx,$$
$$\sigma'_{12} = \int_0^{t_1} E[\Big(z_j - \mu_{z_j}\Big)^2] H(t_1 - x)\bar{G}(x) \exp\{-\Lambda_1(x)\}\lambda_1(x)dx,$$
$$\sigma'_{22} = \int_0^{t_2} E[\Big(z_j - \mu_{z_j}\Big)^2] H(t_2 - x)\bar{G}(x) \exp\{-\Lambda_2(x)\}\lambda_2(x)dx,$$

and

$$\mu_{z_j} = \left[E\left\{z_j H\left(t-x|z_j\right)\overline{G}\left(x|z_j\right)\right\}\right] / \left[E\left\{H\left(t-x|z_j\right)\overline{G}\left(x|z_j\right)\right\}\right].$$

However, in two-stage design, there are two major differences than the design Tsiatis used. One is that there are three treatment groups instead of two. The second is that two different survival endpoints will be used at interim and final analysis. For simplicity, we assume no censoring process ($\bar{G}(x) = 1$) and patient entry patterns in each treatment arm are the same. The covariance terms can be written as

$$\Omega = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \\ \sigma_{12} & \sigma_{22} \end{bmatrix},$$

where

$$\begin{aligned}
\sigma_{11} &= E[\{(z_{1i}+z_{2i})(z_{2i}-\mu_{12})\}^2]\int_0^{t_1} H(t_1-x)\exp\{-\Lambda_1(x)\}\lambda_1(x)dx, \\
\sigma_{12} &= E[\{(z_{1i}+z_{2i})(z_{2i}-\mu_{12})\}\left(z_{ji}+z_{0i}\right)\left(z_{0i}-\mu_{0j}\right)\}] \\
&\quad \times E[\{\triangle_1(t_1)-\Lambda_1(X_1(t_1))\}\{\triangle_2(t_2)-\Lambda_2(X_2(t_2))\}], \\
\sigma_{22} &= E[\{\left(z_{ji}+z_{0i}\right)\left(z_{0i}-\mu_{0j}\right)\}^2]\int_0^{t_2} H(t_2-x)\exp\{-\Lambda_2(x)\}\lambda_2(x)dx. \qquad (10.3)
\end{aligned}$$

The covariance term σ_{12} is derived using the same approach in Tsiatis (1982) [38]. We need to solve

$$E[\{\triangle_1(t_1)-\Lambda_1(X_1(t_1))\}\{\triangle_2(t_2)-\Lambda_2(X_2(t_2))\}], \text{ for } 0 \leqslant t_1 \leqslant t_2,$$

and all other expectation terms.

B. Expectation terms in covariance matrix

Based on Schaid et al [33],

$$S_N^j(t) = \sum_{i=1}^{N} \left(z_{0i} + z_{ji}\right) \Delta_i(t) \left[z_{ji} - \overline{z}_j(t, x_i(t))\right],$$

where N is the total number of patients in the study and

$$\begin{aligned}\overline{z}_j(t,x) &= \sum_{i \in R(t,x)} \left(z_{0i} + z_{ji}\right) z_{ji} / \sum_{i \in R(t,x)} \left(z_{0i} + z_{ji}\right) I\left[X_i(t) \geq x\right] \\ &= \sum_{i \in R(t,x)} z_{ji} / \sum_{i \in R(t,x)} \left(z_{0i} + z_{ji}\right) I\left[X_i(t) \geq x\right].\end{aligned}$$

$R(t,x)$ denotes the risk set for patients in the study. By the Law of Large Number, since

$$\sum_{i \in R(t,x)} z_{ji}/N = \sum_{i=1}^{N} \frac{z_{ji} I\left(X(t) \geq x\right)}{N} \xrightarrow{P} E\left\{z_{j1} I\left(X(t) \geq x\right)\right\},$$

and

$$\begin{aligned}E\left\{z_{j1} I\left(X(t) \geq x\right)\right\} &= E\left\{E\left\{z_{j1} I\left(X(t) \geq x\right) | z_{j1}\right\}\right\} \\ &= E\left\{z_{j1} E\left\{I\left(X(t) \geq x\right) | z_{j1}\right\}\right\} \\ &= E\left\{z_{j1} P\left\{X(t) \geq x | z_{j1}\right\}\right\},\end{aligned}$$

we have

$$\begin{aligned}&\sum_{i \in R(t,x)} z_{ji}/N \xrightarrow{P} E\left\{z_{j1} P\left\{X(t) \geq x | z_{j1}\right\}\right\}, \\ &\sum_{i \in R(t,x)} \left(z_{0i} + z_{ji}\right) I\left(X(t) \geq x\right)/N \xrightarrow{P} E\left\{\left(z_{01} + z_{j1}\right) P\left\{X(t) \geq x | z_{j1}\right\}\right\},\end{aligned}$$

and

$$P\left\{X(t) \geq x\right\} = E\left\{I\left(X(t) \geq x\right)\right\} = E\left\{P\left\{X(t) \geq x | z_{j1}\right\}\right\}$$

$$= \Pr\left(\dot{V} \geq x, t - Y \geq x, W \geq x | z_{j1}\right)$$

$$= \exp\left\{-\Lambda\left(x\right)\right\} E\left\{H\left(t - x | z_{j1}\right) \overline{G}\left(x | z_{j1}\right)\right\}.$$

As $E\left\{z_{j1} H\left(t - x | z_{j1}\right) \overline{G}\left(x | z_{j1}\right)\right\}$ and $E\left\{\left(z_{01} + z_{j1}\right) H\left(t - x | z_{j1}\right) \overline{G}\left(x | z_{j1}\right)\right\}$ are constants,

$$\sum_{i \in R(t,x)} z_{ji} / \sum_{i \in R(t,x)} \left(z_{0i} + z_{ji}\right) I\left(X_i\left(t\right) \geq x\right)$$

converges in distribution to

$$\mu_j\left(t, x\right) = \frac{E\left\{z_{j1} H\left(t - x | z_{j1}\right) \overline{G}\left(x | z_{j1}\right)\right\}}{E\left\{\left(z_{01} + z_{j1}\right) H\left(t - x | z_{j1}\right) \overline{G}\left(x | z_{j1}\right)\right\}}.$$

Because assuming no censoring process and patient entry patterns are identical in each treatment arm. $\mu_j\left(t, x\right)$ will not dependent on time. Thus the notation for $\mu_j\left(t, x\right)$ can be simplified as

$$\mu_j = \frac{E\left(z_{j1}\right)}{E\left(z_{01} + z_{j1}\right)}.$$

So when comparing treatment 2 versus treatment 1 at the interim analysis, we have

$$\mu_{12} = \frac{E\left(z_{21}\right)}{E\left(z_{11} + z_{21}\right)} = \frac{\nu_2}{\nu_1 + \nu_2}.$$

When comparing control versus winning arm at the final analysis, we have

$$\mu_{0j} = \frac{E\left(z_{01}\right)}{E\left(z_{01} + z_{j1}\right)} = \frac{\nu_0}{\nu_0 + \nu_j}.$$

Thus for the covariance matrix in (10.3), the expectation part in σ_{11} is

$$E[\{\left(z_{11} + z_{21}\right)\left(z_{21} - \mu_{12}\right)\}^2]$$

$$= E[\left(z_{11} + z_{21}\right)^2 \left(z_{21} - \mu_{12}\right)^2]$$

$$= E[\left(z_{11} + z_{21}\right)\left(z_{21} - 2 z_{21} \mu_{12} + \mu_{12}^2\right)]$$

$$
\begin{aligned}
&= E[z_{11}\mu_{12}^2 + (z_{21} - 2z_{21}\mu_{12} + z_{21}\mu_{12}^2)] \\
&= E[z_{11}\mu_{12}^2 + z_{21}(1-\mu_{12})^2] \\
&= E[z_{11}]\mu_{12}^2 + E[z_{21}](1-\mu_{12})^2 \\
&= \nu_1\left(\frac{\nu_2}{\nu_1+\nu_2}\right)^2 + \nu_2\left(1-\frac{\nu_2}{\nu_1+\nu_2}\right)^2 \\
&= \frac{\nu_1\nu_2^2}{(\nu_1+\nu_2)^2} + \frac{\nu_1^2\nu_2}{(\nu_1+\nu_2)^2} \\
&= \frac{v_1\nu_2}{\nu_1+\nu_2}.
\end{aligned}
$$

For the expectation part in σ_{22}, we have

$$E[\{(z_{j1}+z_{01})\left(z_{01}-\mu_{0j}\right)\}^2] = \frac{v_0\nu_j}{\nu_0+\nu_j}.$$

For covariance σ_{12}, we have

$$
\begin{aligned}
&E[(z_{j1}+z_{01})\left(z_{01}-\mu_{0j}\right)(z_{11}+z_{21})(z_{21}-\mu_{12})] \\
&= E[z_{j1}\left(z_{01}-\mu_{0j}\right)(z_{11}+z_{21})(z_{21}-\mu_{12})] \\
&= E[z_{j1}\left(z_{01}-\mu_{0j}\right)(z_{11}+z_{21})(z_{21}-\mu_{12})] \\
&= E[z_{j1}(z_{11}+z_{21})\left(z_{01}-\mu_{0j}\right)(z_{21}-\mu_{12})] \\
&= E[z_{j1}\left(z_{01}-\mu_{0j}\right)(z_{21}-\mu_{12})] \\
&= E[-\mu_{0j}z_{j1}(z_{21}-\mu_{12})] \\
&= -\mu_{0j}E[z_{j1}(z_{21}-\mu_{12})]
\end{aligned}
$$

When $j = 1$,

$$
\begin{aligned}
&E[(z_{j1}+z_{01})\left(z_{01}-\mu_{0j}\right)(z_{11}+z_{21})(z_{21}-\mu_{12})] \\
&= -\mu_{01}E\left[z_{11}(z_{21}-\mu_{12})\right]
\end{aligned}
$$

$$= -\mu_{01} E\left[-z_{11}\mu_{12}\right]$$

$$= \mu_{01}\mu_{12} E\left[z_{11}\right]$$

$$= \frac{v_0}{\nu_0+\nu_1}\frac{\nu_2}{\nu_1+\nu_2} v_1$$

$$= \frac{v_0\nu_1\nu_2}{(\nu_0+\nu_1)(\nu_1+\nu_2)},$$

and when $j = 2$

$$E[(z_{j1}+z_{01})\left(z_{01}-\mu_{0j}\right)(z_{11}+z_{21})(z_{21}-\mu_{12})]$$

$$= -\mu_{02} E\left[z_{21}(z_{21}-\mu_{12})\right]$$

$$= -\mu_{01} E\left[z_{21}(1-\mu_{12})\right]$$

$$= -\mu_{01}(1-\mu_{12}) E\left[z_{21}\right]$$

$$= \frac{-v_0}{\nu_0+\nu_1}\left(1-\frac{\nu_2}{\nu_1+\nu_2}\right) v_2$$

$$= \frac{-v_0\nu_1\nu_2}{(\nu_0+\nu_1)(\nu_1+\nu_2)}.$$

Thus

$$E[(z_{j1}+z_{01})\left(z_{01}-\mu_{0j}\right)(z_{11}+z_{21})(z_{21}-\mu_{12})]$$

$$= (-1)^{j-1}\frac{v_0\nu_1\nu_2}{(\nu_0+\nu_1)(\nu_1+\nu_2)}.$$

C. Covariance of log-rank test statistics in the case of bivariate exponential distribution

Suppose $\theta_0 \sim Exp(\lambda_0)$, $\theta_2 \sim Exp(\lambda_2)$. Let $\theta_1 \sim \min(\theta_0, \theta_2)$. Then $\theta_1 \sim Exp(\lambda_1)$, with $\lambda_1 = \lambda_0 + \lambda_2$. It assumed that θ_0 is the time of PFS , θ_2 is the time of death, and θ_1 is the recorded time of PFS. Let Y denote the real time of entry. We will first consider everything conditional on Y, that is, treat Y as fixed.

Similar in Tsiatis [38], denote

$$X_1(t_1) = \max(\min(\theta_1, t_1 - Y), 0)$$

$$X_2(t_2) = \max(\min(\theta_2, t_2 - Y), 0)$$

$$\Delta_1(t_1) = 1 \text{ iff } \theta_1 < t_1 - Y$$

$$\Delta_2(t_2) = 1 \text{ iff } \theta_2 < t_2 - Y.$$

We want to show

Theorem

$$E\{[\Delta_1(t_1) - \lambda_1 X_1(t_1)][\Delta_2(t_2) - \lambda_2 X_2(t_2)]\} = \frac{\lambda_2}{\lambda_1}\left(1 - e^{-\lambda_1(t_1 - Y)}\right), \text{ for } 0 \leqslant t_1 \leqslant t_2.$$

Lemma 1:

$$E\{[\Delta_1(t_1) - \lambda_1 X_1(t_1)][\Delta_2(t_2) - \lambda_2 X_2(t_2)]\}$$

$$= E\{[\Delta_1(t_1) - \lambda_1 X_1(t_1)][\Delta_2(t_1) - \lambda_2 X_2(t_1)]\}, \; for \; 0 \leqslant t_1 \leqslant t_2$$

i.e

$$E\{[\Delta_1(t_1) - \lambda_1 X_1(t_1)][\Delta_2(t_2) - \lambda_2 X_2(t_2) - (\Delta_2(t_1) - \lambda_2 X_2(t_1))]\} = 0$$

Proof: We consider two cases: $\theta_2 > \theta_0$ and $\theta_2 \leq \theta_0$. Denote the integration on the first set as A and the integration on the second set as B. Thus we have

$$A = E\left\{I_{\{\theta_2 > \theta_0\}}[\Delta_1(t_1) - \lambda_1 X_1(t_1)][\Delta_2(t_2) - \lambda_2 X_2(t_2) - (\Delta_2(t_1) - \lambda_2 X_2(t_1))]\right\}$$

$$= E\left\{E\left\{I_{\{\theta_2 > \theta_0\}}[\Delta_1(t_1) - \lambda_1 X_1(t_1)]|\theta_2\right\}[\Delta_2(t_2) - \lambda_2 X_2(t_2) - (\Delta_2(t_1) - \lambda_2 X_2(t_1))]\right\}$$

and

$$B = E\left\{I_{\{\theta_2 \leq \theta_0\}}[\Delta_1(t_1) - \lambda_1 X_1(t_1)][\Delta_2(t_2) - \lambda_2 X_2(t_2) - (\Delta_2(t_1) - \lambda_2 X_2(t_1))]\right\}$$

$$= E\left\{E\left\{I_{\{\theta_2 \leq \theta_0\}}\left[\Delta_1(t_1) - \lambda_1 X_1(t_1)\,|\theta_2\right]\right\}\left[\Delta_2(t_2) - \lambda_2 X_2(t_2) - (\Delta_2(t_1) - \lambda_2 X_2(t_1))\right]\right\}.$$

For the first set, consider the conditional expectation on $\theta_2 (\geqslant t_1 - Y)$,

$$\begin{aligned}
&E\left\{I_{\{\theta_2>\theta_0\}}\left[\Delta_1(t_1) - \lambda_1 X_1(t_1)\right]|\theta_2\right\} \\
&= E\left\{I_{\{\theta_2>\theta_0\}}\left[I_{\{\theta_0<t_1-Y\}} - \lambda_1\theta_0 I_{\{\theta_0<t_1-Y\}} - \lambda_1(t_1-Y)\,I_{(\theta_0\geq t_1-Y)}\right]|\theta_2\right\} \\
&= E\left\{I_{\{\theta_0<t_1-Y\}} - \lambda_1\theta_0 I_{\{\theta_0<t_1-Y\}} - \lambda_1(t_1-Y)\,I_{(\theta_2>\theta_0\geq t_1-Y)}|\theta_2\right\} \\
&= \int_0^{t_1-Y}\lambda_0 e^{-\lambda_0\theta_0}d\theta_0 - \int_0^{t_1-Y}\lambda_1\theta_0\lambda_0 e^{-\lambda_0\theta_0}d\theta_0 - \lambda_1(t_1-Y)\int_{t_1-Y}^{\theta_2}\lambda_0 e^{-\lambda_0\theta_0}d\theta_0 \\
&= 1 - e^{-\lambda_0(t_1-Y)} + \lambda_1\int_0^{t_1-Y}\theta_0 de^{-\lambda_0\theta_0} - \lambda_1(t_1-Y)\left(e^{-\lambda_0(t_1-Y)} - e^{-\lambda_0\theta_2}\right) \\
&= 1 - e^{-\lambda_0(t_1-Y)} + \lambda_1(t_1-Y)e^{-\lambda_0(t_1-Y)} - \lambda_1\int_0^{t_1-Y}e^{-\lambda_0\theta_0}d\theta_0 - \lambda_1(t_1-Y)\left(e^{-\lambda_0(t_1-Y)} - e^{-\lambda_0\theta_2}\right) \\
&= 1 - e^{-\lambda_0(t_1-Y)} + \lambda_1(t_1-Y)e^{-\lambda_0\theta_2} - \frac{\lambda_1}{\lambda_0}\left(1 - e^{-\lambda_0(t_1-Y)}\right) \\
&= \lambda_1(t_1-Y)e^{-\lambda_0\theta_2} - \frac{\lambda_2}{\lambda_0}\left(1 - e^{-\lambda_0(t_1-Y)}\right).
\end{aligned}$$

Therefore

$$\begin{aligned}
A &= E\left\{\left[\lambda_1(t_1-Y)e^{-\lambda_0\theta_2} - \frac{\lambda_2}{\lambda_0}\left(1 - e^{-\lambda_0(t_1-Y)}\right)\right]\left[\Delta_2(t_2) - \lambda_2 X(t_2) - (\Delta_2(t_1) - \lambda_2 X(t_1))\right]\right\} \\
&= E\left\{\lambda_1(t_1-Y)e^{-\lambda_0\theta_2}\left[\Delta_2(t_2) - \lambda_2 X_2(t_2) - (\Delta_2(t_1) - \lambda_2 X_2(t_1))\right]\right\}.
\end{aligned}$$

In the above, notice that

$$\begin{aligned}
&\Delta_2(t_2) - \lambda_2 X_2(t_2) - (\Delta_2(t_1) - \lambda_2 X_2(t_1)) \\
&= I_{\{\theta_2<t_2-Y\}} - \lambda_2\theta_2 I_{\{\theta_2<t_2-Y\}} - \lambda_2(t_2-Y)\,I_{\{\theta_2\geq t_2-Y\}} \\
&\quad - I_{\{\theta_2<t_1-Y\}} + \lambda_2\theta_2 I_{\{\theta_2<t_1-Y\}} + \lambda_2(t_1-Y)\,I_{\{\theta_2\geq t_1-Y\}} \\
&= I_{\{t_1-Y\leq\theta_2<t_2-Y\}} - \lambda_2\theta_2 I_{\{t_1-Y\leq\theta_2<t_2-Y\}} + \lambda_2(t_1-Y)\,I_{\{t_1-Y\leq\theta_2<t_2-Y\}} \\
&\quad - \lambda_2(t_2-t_1)\,I_{\{\theta_2\geq t_2-Y\}}
\end{aligned}$$

$$= I_{\{t_1-Y\leq\theta_2<t_2-Y\}} - \lambda_2\left(\theta_2-(t_1-Y)\right) I_{\{t_1-Y\leq\theta_2<t_2-Y\}} - \lambda_2\left(t_2-t_1\right) I_{\{\theta_2\geq t_2-Y\}}$$

is not zero only when $\theta_2 \geqslant t_1 - Y$.

Now let's look at set B. Consider the conditional expectation on $\theta_2\left(\geqslant t_1-Y\right)$.

$$\begin{aligned}
&E\left\{I_{\{\theta_2\leq\theta_0\}}\left[\Delta_1\left(t_1\right)-\lambda_1 X_1\left(t_1\right)|\theta_2\right]\right\} \\
&= E\left\{I_{\{\theta_2\leq\theta_0\}}\left[I_{\{\theta_2<t_1-Y\}}-\lambda_1\theta_2 I_{\{\theta_2<t_1-Y\}}-\lambda_1\left(t_1-Y\right)I_{(\theta_2\geq t_1-Y)}|\theta_2\right]\right\} \\
&= E\left\{-I_{\{\theta_2\leq\theta_0\}}\lambda_1\left(t_1-Y\right)I_{(\theta_2\geq t_1-Y)}|\theta_2\right\} \\
&= -e^{-\lambda_0\theta_2}\lambda_1\left(t_1-Y\right)I_{(\theta_2\geq t_1-Y)} \\
&= -\lambda_1\left(t_1-Y\right)e^{-\lambda_0\theta_2}I_{(\theta_2\geq t_1-Y)}.
\end{aligned}$$

Hence

$$B = -E\left\{\lambda_1\left(t_1-Y\right)e^{-\lambda_0\theta_2}\left[\Delta_2\left(t_2\right)-\lambda_2 X_2\left(t_2\right)-\left(\Delta_2\left(t_1\right)-\lambda_2 X_2\left(t_1\right)\right)\right]\right\}$$

and

$$\begin{aligned}
&E\left\{\left[\Delta_1\left(t_1\right)-\lambda_1 X_1\left(t_1\right)\right]\left[\Delta_2\left(t_2\right)-\lambda_2 X_2\left(t_2\right)-\left(\Delta_2\left(t_1\right)-\lambda_2 X_2\left(t_1\right)\right)\right]\right\} \\
&= A + B = 0.
\end{aligned}$$

That finishes the proof of lemma 1. The covariance of PFS part with the increment of OS part is zero.

Proof of theorem

Again we divide $E\left\{\left[\Delta_1\left(t\right)-\lambda_1 X_1\left(t\right)\right]\left[\Delta_2\left(t\right)-\lambda_2 X_2\left(t\right)\right]\right\}$ into two parts.

$$A = E\left\{I_{\{\theta_2>\theta_0\}}\left[\Delta_1\left(t\right)-\lambda_1 X_1\left(t\right)\right]\left[\Delta_2\left(t\right)-\lambda_2 X_2\left(t\right)\right]\right\},$$

$$B = E\left\{I_{\{\theta_2\leq\theta_0\}}\left[\Delta_1\left(t\right)-\lambda_1 X_1\left(t\right)\right]\left[\Delta_2\left(t\right)-\lambda_2 X_2\left(t\right)\right]\right\}.$$

$$
\begin{aligned}
B &= E\left\{I_{\{\theta_2\leq\theta_0\}}\left[\Delta_1(t)-\lambda_1 X_1(t)\right]\left[\Delta_2(t)-\lambda_2 X_2(t)\right]\right\} \\
&= E\left\{I_{\{\theta_2\leq\theta_0\}}\left[I_{\{\theta_2<t-Y\}}-\lambda_1\theta_2 I_{\{\theta_2<t-Y\}}-\lambda_1(t-Y)I_{(\theta_2\geq t-Y)}\right]\right. \\
&\qquad \left.\times\left[I_{\{\theta_2<t-Y\}}-\lambda_2\theta_2 I_{\{\theta_2<t-Y\}}-\lambda_2(t-Y)I_{(\theta_2\geq t-Y)}\right]\right\} \\
&= E\left\{e^{-\lambda_0\theta_2}\left[I_{\{\theta_2<t-Y\}}-(\lambda_1+\lambda_2)\theta_2 I_{\{\theta_2<t-Y\}}+\lambda_1\lambda_2\theta_2^2 I_{\{\theta_2<t-Y\}}+\lambda_1\lambda_2(t-Y)^2 I_{(\theta_2\geq t-Y)}\right]\right\} \\
&= \int_0^{t-Y}\lambda_2 e^{-\lambda_2\theta_2}e^{-\lambda_0\theta_2}d\theta_2-(\lambda_1+\lambda_2)\int_0^{t-Y}\lambda_2 e^{-\lambda_2\theta_2}\theta_2 e^{-\lambda_0\theta_2}d\theta_2 \\
&\qquad +\lambda_1\lambda_2\int_0^{t-Y}\lambda_2 e^{-\lambda_2\theta_2}\theta_2^2 e^{-\lambda_0\theta_2}d\theta_2+\lambda_1\lambda_2(t-Y)^2\int_{t-Y}^{\infty}\lambda_2 e^{-\lambda_2\theta_2}e^{-\lambda_0\theta_2}d\theta_2 \\
&= \frac{\lambda_2}{\lambda_0+\lambda_2}\left(1-e^{-(\lambda_0+\lambda_2)(t-Y)}\right)+\lambda_1\lambda_2(t-Y)^2\frac{\lambda_2}{\lambda_0+\lambda_2}e^{-(\lambda_0+\lambda_2)(t-Y)} \\
&\qquad -\frac{\lambda_2(\lambda_1+\lambda_2)}{\lambda_0+\lambda_2}\left[\frac{1}{\lambda_0+\lambda_2}\left(1-e^{-(\lambda_0+\lambda_2)(t-Y)}\right)-(t-Y)e^{-(\lambda_0+\lambda_2)(t-Y)}\right] \\
&\qquad +\frac{\lambda_1\lambda_2^2}{\lambda_0+\lambda_2}\left[-(t-Y)^2 e^{-(\lambda_0+\lambda_2)(t-Y)}+\frac{2}{(\lambda_0+\lambda_2)^2}\left(1-e^{-(\lambda_0+\lambda_2)(t-Y)}\right)\right. \\
&\qquad \left.-(t-Y)\frac{2}{\lambda_0+\lambda_2}e^{-(\lambda_0+\lambda_2)(t-Y)}\right] \\
&= \left[\frac{\lambda_2}{\lambda_0+\lambda_2}-\frac{\lambda_2(\lambda_1+\lambda_2)}{(\lambda_0+\lambda_2)^2}+\frac{2\lambda_1\lambda_2^2}{(\lambda_0+\lambda_2)^3}\right]\left(1-e^{-(\lambda_0+\lambda_2)(t-Y)}\right) \\
&\qquad +\left[\frac{\lambda_2(\lambda_1+\lambda_2)}{\lambda_0+\lambda_2}-\frac{2\lambda_1\lambda_2^2}{(\lambda_0+\lambda_2)^2}\right](t-Y)e^{-(\lambda_0+\lambda_2)(t-Y)} \\
&\qquad +\left(\frac{\lambda_1\lambda_2^2}{\lambda_0+\lambda_2}-\frac{\lambda_1\lambda_2^2}{\lambda_0+\lambda_2}\right)(t-Y)^2 e^{-(\lambda_0+\lambda_2)(t-Y)} \\
&= \frac{\lambda_2^2}{(\lambda_0+\lambda_2)^2}\left(1-e^{-(\lambda_0+\lambda_2)(t-Y)}\right)+\frac{\lambda_0\lambda_2}{\lambda_0+\lambda_2}(t-Y)e^{-(\lambda_0+\lambda_2)(t-Y)}
\end{aligned}
$$

As for A, we calculate it as follow:

$$
\begin{aligned}
A &= E\left\{I_{\{\theta_2>\theta_0\}}\left[\Delta_1(t)-\lambda_1 X_1(t)\right]\left[\Delta_2(t)-\lambda_2 X_2(t)\right]\right\} \\
&= E\left\{I_{\{\theta_2>\theta_0\}}\left[I_{\{\theta_0<t-Y\}}-\lambda_1\theta_0 I_{\{\theta_0<t-Y\}}-\lambda_1(t-Y)I_{(\theta_0\geq t-Y)}\right]\right. \\
&\qquad \left.\times\left[I_{\{\theta_2<t-Y\}}-\lambda_2\theta_2 I_{\{\theta_2<t-Y\}}-\lambda_2(t-Y)I_{(\theta_2\geq t-Y)}\right]\right\}
\end{aligned}
$$

$$= E\left\{\left[I_{\{\theta_0<t-Y\}} - \lambda_1\theta_0 I_{\{\theta_0<t-Y\}} - \lambda_1\left(t-Y\right)I_{(\theta_0\geq t-Y)}\right]\right.$$
$$\left.\times E\left\{I_{\{\theta_2>\theta_0\}}\left[I_{\{\theta_2<t-Y\}} - \lambda_2\theta_2 I_{\{\theta_2<t-Y\}} - \lambda_2\left(t-Y\right)I_{(\theta_2\geq t-Y)}\right]|\theta_0\right\}\right\}.$$

To calculate the conditional expectation,

$$E\left\{I_{\{\theta_2>\theta_0\}}\left[I_{\{\theta_2<t-Y\}} - \lambda_2\theta_2 I_{\{\theta_2<t-Y\}} - \lambda_2\left(t-Y\right)I_{(\theta_2\geq t-Y)}\right]|\theta_0\right\}$$
$$= E\left\{I_{\{\theta_0<\theta_2<t-Y\}} - \lambda_2\theta_2 I_{\{\theta_0<\theta_2<t-Y\}} - \lambda_2\left(t-Y\right)I_{(\theta_2\geq(t-Y)\vee\theta_0)}|\theta_0\right\}$$
$$= I_{\{\theta_0<t-Y\}}\left(e^{-\lambda_0\theta_0} - e^{-\lambda_2(t-Y)}\right) - I_{\{\theta_0<t-Y\}}\lambda_2\int_{\theta_0}^{t-Y}\theta_2\lambda_2 e^{-\lambda_2\theta_2}d\theta_2 - \lambda_2\left(t-Y\right)e^{-\lambda_2(t-Y)\vee\theta_0}$$
$$= I_{\{\theta_0<t-Y\}}\left(e^{-\lambda_0\theta_0} - e^{-\lambda_2(t-Y)}\right)$$
$$- I_{\{\theta_0<t-Y\}}\lambda_2\left[\frac{1}{\lambda_2}\left(e^{-\lambda_2\theta_0} - e^{-\lambda_2(t-Y)}\right) + \theta_0 e^{-\lambda_2\theta_0} - \left(t-Y\right)e^{-\lambda_2(t-Y)}\right]$$
$$- \lambda_2\left(t-Y\right)e^{-\lambda_2(t-Y)\vee\theta_0}$$
$$= -I_{\{\theta_0<t-Y\}}\lambda_2\theta_0 e^{-\lambda_2\theta_0} + I_{\{\theta_0<t-Y\}}\lambda_2\left(t-Y\right)e^{-\lambda_2(t-Y)} - \lambda_2\left(t-Y\right)e^{-\lambda_2(t-Y)\vee\theta_0}$$
$$= -\lambda_2\left(\left(t-Y\right)\wedge\theta_0\right)e^{-\lambda_2\theta_0}$$

That implies that,

$$A = E\left\{\left[I_{\{\theta_0<t-Y\}} - \lambda_1\theta_0 I_{\{\theta_0<t-Y\}} - \lambda_1\left(t-Y\right)I_{(\theta_0\geqslant t-Y)}\right]\left[-\lambda_2\left(\left(t-Y\right)\wedge\theta_0\right)e^{-\lambda_2\theta_0}\right]\right\}$$
$$= E\left\{-\lambda_2\theta_0 e^{-\lambda_2\theta_0}I_{\{\theta_0<t-Y\}} + \lambda_1\lambda_2\theta_0^2 e^{-\lambda_2\theta_0}I_{\{\theta_0<t-Y\}} + \lambda_1\lambda_2\left(t-Y\right)^2 e^{-\lambda_2\theta_0}I_{(\theta_0\geqslant t-Y)}\right\}$$
$$= -\lambda_2\int_0^{t-Y}\theta_0 e^{-\lambda_2\theta_0}\lambda_0 e^{-\lambda_0\theta_0}d\theta_0 + \lambda_1\lambda_2\int_0^{t-Y}\theta_0^2 e^{-\lambda_2\theta_0}\lambda_0 e^{-\lambda_0\theta_0}d\theta_0$$
$$+ \lambda_1\lambda_2\left(t-Y\right)^2\int_{t-Y}^{\infty}e^{-\lambda_2\theta_0}\lambda_0 e^{-\lambda_0\theta_0}d\theta_0$$
$$= -\frac{\lambda_0\lambda_2}{\lambda_0+\lambda_2}\left(\frac{1}{\lambda_0+\lambda_2}\left(1-e^{-(\lambda_0+\lambda_2)(t-Y)}\right) - \left(t-Y\right)e^{-(\lambda_0+\lambda_2)(t-Y)}\right)$$
$$+ \lambda_1\lambda_2\left(t-Y\right)^2\frac{\lambda_0}{\lambda_0+\lambda_2}e^{-(\lambda_0+\lambda_2)(t-Y)}$$
$$+ \frac{\lambda_0\lambda_1\lambda_2}{\lambda_0+\lambda_2}\left[-\left(t-Y\right)^2 e^{-(\lambda_0+\lambda_2)(t-Y)}\right.$$

$$
\begin{aligned}
&+\frac{2}{\lambda_0+\lambda_2}\left(\frac{1}{\lambda_0+\lambda_2}\left(1-e^{-(\lambda_0+\lambda_2)(t-Y)}\right)-(t-Y)\,e^{-(\lambda_0+\lambda_2)(t-Y)}\right)\Bigg] \\
&=\left(1-e^{-(\lambda_0+\lambda_2)(t-Y)}\right)\left[-\frac{\lambda_0\lambda_2}{(\lambda_0+\lambda_2)^2}+\frac{2\lambda_0\lambda_2}{(\lambda_0+\lambda_2)^2}\right] \\
&\quad+(t-Y)\,e^{-(\lambda_0+\lambda_2)(t-Y)}\left[\frac{\lambda_0\lambda_2}{\lambda_0+\lambda_2}-\frac{2\lambda_0\lambda_2}{\lambda_0+\lambda_2}\right] \\
&\quad+(t-Y)^2\,e^{-(\lambda_0+\lambda_2)(t-Y)}\left[\lambda_0\lambda_2-\lambda_0\lambda_2\right] \\
&=\left(1-e^{-(\lambda_0+\lambda_2)(t-Y)}\right)\frac{\lambda_0\lambda_2}{(\lambda_0+\lambda_2)^2}-(t-Y)\,e^{-(\lambda_0+\lambda_2)(t-Y)}\frac{\lambda_0\lambda_2}{\lambda_0+\lambda_2}.
\end{aligned}
$$

Therefore, from Lemma 1, the theorem is proved. Since

$$
\begin{aligned}
&E\left\{\left[\Delta_1(t)-\lambda_1X_1(t)\right]\left[\Delta_2(t)-\lambda_2X_2(t)\right]\right\} \\
&=A+B \\
&=\left(1-e^{-(\lambda_0+\lambda_2)(t-Y)}\right)\left[\frac{\lambda_0\lambda_2}{(\lambda_0+\lambda_2)^2}+\frac{\lambda_2^2}{(\lambda_0+\lambda_2)^2}\right] \\
&\quad-(t-Y)\,e^{-(\lambda_0+\lambda_2)(t-Y)}\left[\frac{\lambda_0\lambda_2}{\lambda_0+\lambda_2}-\frac{\lambda_0\lambda_2}{\lambda_0+\lambda_2}\right] \\
&=\frac{\lambda_2}{\lambda_0+\lambda_2}\left(1-e^{-(\lambda_0+\lambda_2)(t-Y)}\right) \\
&=\frac{\lambda_2}{\lambda_1}\left(1-e^{-\lambda_1(t-Y)}\right).
\end{aligned}
$$

In addition, we can show that

Lemma 2:

$$
\begin{aligned}
&E\left\{\left[\Delta_1(t_2)-\lambda_1X_1(t_2)\right]\left[\Delta_2(t_1)-\lambda_2X_2(t_1)\right]\right\} \\
&=E\left\{\left[\Delta_1(t_1)-\lambda_1X_1(t_1)\right]\left[\Delta_2(t_1)-\lambda_2X_2(t_1)\right]\right\},\ \text{for } 0\le t_2\le t_1.
\end{aligned}
$$

Proof. For lemma 2, we only need to show that

$$
E\left\{\left[\Delta_1(t_2)-\lambda_1X_1(t_2)-\left(\Delta_1(t_1)-\lambda_1X_1(t_1)\right)\right]\left[\Delta_2(t_1)-\lambda_2X_2(t_1)\right]\right\}=0.
$$

Similarly, we let

$$A = E\left\{ I_{[\theta_2>\theta_0]} \left[\Delta_1(t_2) - \lambda_1 X_1(t_2) - (\Delta_1(t_1) - \lambda_1 X_1(t_1))\right] \left[\Delta_2(t_1) - \lambda_2 X_2(t_1)\right] \right\},$$

$$B = E\left\{ I_{[\theta_2\leq\theta_0]} \left[\Delta_1(t_2) - \lambda_1 X_1(t_2) - (\Delta_1(t_1) - \lambda_1 X_1(t_1))\right] \left[\Delta_2(t_1) - \lambda_2 X_2(t_1)\right] \right\}.$$

On the set $\theta_2 > \theta_0$,

$$\begin{aligned}
&\Delta_1(t_2) - \lambda_1 X_1(t_2) - (\Delta_1(t_1) - \lambda_1 X_1(t_1)) \\
&= I_{\{\theta_0<t_2-Y\}} - \lambda_1\theta_0 I_{\{\theta_0<t_2-Y\}} - \lambda_1(t_2 - Y) I_{(\theta_0\geqslant t_2-Y)} \\
&\qquad - I_{\{\theta_0<t_1-Y\}} + \lambda_1\theta_0 I_{\{\theta_0<t_1-Y\}} + \lambda_1(t_1 - Y) I_{(\theta_0\geqslant t_1-Y)} \\
&= I_{\{t_1-Y\leq\theta_0<t_2-Y\}} - \lambda_1\theta_0 I_{\{t_1-Y\leq\theta_0<t_2-Y\}} + \lambda_1(t_1 - Y) I_{\{t_1-Y\leq\theta_0<t_2-Y\}} \\
&\qquad - \lambda_1(t_2 - t_1) I_{(\theta_0\geqslant t_2-Y)} \\
&= I_{\{t_1-Y\leq\theta_0<t_2-Y\}} - \lambda_1(\theta_0 - (t_1 - Y)) I_{\{t_1-Y\leq\theta_0<t_2-Y\}} - \lambda_1(t_2 - t_1) I_{(\theta_0\geqslant t_2-Y)}
\end{aligned}$$

This is not zero only on $\theta_0 \geqslant t_1 - Y$. Therefore, we consider the conditional expectation on $\theta_0 (\geqslant t_1 - Y)$.

$$\begin{aligned}
&E\left\{ I_{\{\theta_2>\theta_0\}} (\Delta_2(t_1) - \lambda_2 X_2(t_1)) \,|\theta_0 \right\} \\
&= E\left\{ I_{\{\theta_2>\theta_0\}} \left[I_{\{\theta_2<t_1-Y\}} - \lambda_2\theta_2 I_{\{\theta_2<t_1-Y\}} - \lambda_2(t_1 - Y) I_{\{\theta_2\geq t_1-Y\}} \right] |\theta_0 \right\} \\
&= E\left\{ -\lambda_2(t_1 - Y) I_{\{\theta_2>\theta_0\}} \right\} \\
&= -\lambda_2(t_1 - Y) e^{-\lambda_2\theta_0}.
\end{aligned}$$

$$\begin{aligned}
A &= E\left\{ I_{[\theta_2>\theta_0]} \left[\Delta_1(t_2) - \lambda_1 X_1(t_2) - (\Delta_1(t_1) - \lambda_1 X_1(t_1))\right] \left[\Delta_2(t_1) - \lambda_2 X_2(t_1)\right] \right\} \\
&= E\left\{ I_{\{t_1-Y\leq\theta_0<t_2-Y\}} - \lambda_1(\theta_0 - (t_1 - Y)) I_{\{t_1-Y\leq\theta_0<t_2-Y\}} - \lambda_1(t_2 - t_1) I_{(\theta_0\geqslant t_2-Y)} \right. \\
&\qquad \left. \times \left[-\lambda_2(t_1 - Y) e^{-\lambda_2\theta_0} \right] \right.
\end{aligned}$$

$$
\begin{aligned}
&= -\lambda_2 (t_1 - Y) \int_{t_1-Y}^{t_2-Y} e^{-\lambda_2\theta_0}\lambda_0 e^{-\lambda_0\theta_0} d\theta_0 \\
&\quad + \lambda_1\lambda_2 (t_1 - Y) \int_{t_1-Y}^{t_2-Y} (\theta_0 - (t_1 - Y)) e^{-\lambda_2\theta_0}\lambda_0 e^{-\lambda_0\theta_0} d\theta_0 \\
&\quad + \lambda_1\lambda_2 (t_2 - t_1)(t_1 - Y) \int_{t_2-Y}^{\infty} e^{-\lambda_2\theta_0}\lambda_0 e^{-\lambda_0\theta_0} d\theta_0 \\
&= -\lambda_2 (t_1 - Y) \frac{\lambda_0}{\lambda_0+\lambda_2} \left[e^{-(\lambda_0+\lambda_2)(t_1-Y)} - e^{-(\lambda_0+\lambda_2)(t_2-Y)}\right] \\
&\quad + \frac{\lambda_0\lambda_1\lambda_2 (t_1 - Y)}{\lambda_0+\lambda_2} \left[-(t_2 - t_1) e^{-(\lambda_0+\lambda_2)(t_1-Y)} + \right. \\
&\quad \left. + \frac{1}{\lambda_0+\lambda_2}\left(e^{-(\lambda_0+\lambda_2)(t_1-Y)} - e^{-(\lambda_0+\lambda_2)(t_2-Y)}\right)\right] \\
&\quad + \frac{\lambda_0\lambda_1\lambda_2 (t_2 - t_1)(t_1 - Y)}{\lambda_0+\lambda_2} e^{-(\lambda_0+\lambda_2)(t_2-Y)} \\
&= 0.
\end{aligned}
$$

On the set $\theta_2 \leq \theta_0$,

$$
\begin{aligned}
&\Delta_1 (t_2) - \lambda_1 X_1 (t_2) - (\Delta_1 (t_1) - \lambda_1 X_1 (t_1)) \\
&= I_{\{\theta_2 < t_2 - Y\}} - \lambda_1\theta_2 I_{\{\theta_2 < t_2 - Y\}} - \lambda_1 (t_2 - Y) I_{(\theta_2 \geqslant t_2 - Y)} \\
&\quad - I_{\{\theta_2 < t_1 - Y\}} + \lambda_1\theta_2 I_{\{\theta_2 < t_1 - Y\}} + \lambda_1 (t_1 - Y) I_{(\theta_2 \geqslant t_1 - Y)} \\
&= I_{\{t_1 - Y \leq \theta_2 < t_2 - Y\}} - \lambda_1 (\theta_2 - (t_1 - Y)) I_{\{t_1 - Y \leq \theta_2 < t_2 - Y\}} - \lambda_1 (t_2 - t_1) I_{(\theta_2 \geqslant t_2 - Y)}.
\end{aligned}
$$

We consider the conditional expectation on $\theta_2 (\geqslant t_1 - Y)$.

$$
\begin{aligned}
&E\left\{I_{\{\theta_2 \leq \theta_0\}} (\Delta_2 (t_1) - \lambda_2 X_2 (t_1)) | \theta_2\right\} \\
&= E\left\{I_{\{\theta_2 \leq \theta_0\}} \left[I_{\{\theta_2 < t_1 - Y\}} - \lambda_2\theta_2 I_{\{\theta_2 < t_1 - Y\}} - \lambda_2 (t_1 - Y) I_{\{\theta_2 \geq t_1 - Y\}}\right] | \theta_2\right\} \\
&= -\lambda_2 (t_1 - Y) I_{\{\theta_2 \geq t_1 - Y\}} e^{-\lambda_0\theta_2}.
\end{aligned}
$$

$$
B = E\left\{I_{[\theta_2 \leq \theta_0]} [\Delta_1 (t_2) - \lambda_1 X_1 (t_2) - (\Delta_1 (t_1) - \lambda_1 X_1 (t_1))] [\Delta_2 (t_1) - \lambda_2 X_2 (t_1)]\right\}
$$

$$
\begin{aligned}
&= E\left\{ I_{\{t_1-Y\le\theta_2<t_2-Y\}} - \lambda_1\left(\theta_2-(t_1-Y)\right) I_{\{t_1-Y\le\theta_2<t_2-Y\}} - \lambda_1\left(t_2-t_1\right) I_{(\theta_2\ge t_2-Y)} \right. \\
&\qquad \left. \times \left[-\lambda_2\left(t_1-Y\right) I_{(\theta_2\ge t_2-Y)} e^{-\lambda_0\theta_2} \right] \right. \\
&= -\lambda_2\left(t_1-Y\right)\int_{t_1-Y}^{t_2-Y} e^{-\lambda_0\theta_2}\lambda_2 e^{-\lambda_2\theta_2} d\theta_2 \\
&\qquad + \lambda_1\lambda_2\left(t_1-Y\right)\int_{t_1-Y}^{t_2-Y}\left(\theta_2-(t_1-Y)\right) e^{-\lambda_0\theta_2}\lambda_2 e^{-\lambda_2\theta_2} d\theta_2 \\
&\qquad + \lambda_1\lambda_2\left(t_2-t_1\right)\left(t_1-Y\right)\int_{t_2-Y}^{\infty} e^{-\lambda_0\theta_2}\lambda_2 e^{-\lambda_2\theta_2} d\theta_2 \\
&= -\lambda_2\left(t_1-Y\right)\frac{\lambda_2}{\lambda_0+\lambda_2}\left[e^{-(\lambda_0+\lambda_2)(t_1-Y)} - e^{-(\lambda_0+\lambda_2)(t_2-Y)} \right] \\
&\qquad + \frac{\lambda_1\lambda_2^2\left(t_1-Y\right)}{\lambda_0+\lambda_2}\left[-\left(t_2-t_1\right) e^{-(\lambda_0+\lambda_2)(t_1-Y)} + \frac{1}{\lambda_0+\lambda_2}\left(e^{-(\lambda_0+\lambda_2)(t_1-Y)} - e^{-(\lambda_0+\lambda_2)(t_2-Y)} \right) \right] \\
&\qquad + \frac{\lambda_1\lambda_2^2\left(t_2-t_1\right)\left(t_1-Y\right)}{\lambda_0+\lambda_2} e^{-(\lambda_0+\lambda_2)(t_2-Y)} \\
&= 0
\end{aligned}
$$

So $E\left\{\left[\Delta_1\left(t_2\right)-\lambda_1 X_1\left(t_2\right)-\left(\Delta_1\left(t_1\right)-\lambda_1 X_1\left(t_1\right)\right)\right]\left[\Delta_2\left(t_1\right)-\lambda_2 X_2\left(t_1\right)\right]\right\} = A + B = 0.$

Based on the derivation above, the covariance term is

$$
\begin{aligned}
&E[\left\{\left(z_{1i}+z_{2i}\right)\left(z_{2i}-\mu_{12}\left(t,x\right)\right)\right\}\left(z_{ji}+z_{0i}\right)\left(z_{0i}-\mu_{0j}\left(t,x\right)\right)\}] \\
&\qquad \times E[\{\triangle_1(t_1)-\Lambda_1(X_1(t_1))\}\{\triangle_2(t_2)-\Lambda_2(X_2(t_2))\}] \\
&\qquad = \frac{v_0\nu_1\nu_2}{\left(\nu_0+\nu_1\right)\left(\nu_1+\nu_2\right)}\left(\frac{\lambda_2}{\lambda_1}\left(1-e^{-\lambda_1(t_1-Y)}\right)\right),\ for\ 0\leqslant t_1\leqslant t_2.
\end{aligned}
$$

Similarly the variances can be calculated as

$$
E[\left\{\left(z_{1i}+z_{2i}\right)\left(z_{2i}-\mu_{12}\left(t,x\right)\right)\right\}^2]E\left[\left(\Delta_1\left(t_1\right)-\Lambda_1\left(X_1\left(t_1\right)\right)\right)^2\right] = \frac{v_1\nu_2}{\nu_1+\nu_2}\left(1-e^{-\lambda_1(t_1-Y)}\right)
$$

and

$$
E[\left\{\left(z_{ji}+z_{0i}\right)\left(z_{0i}-\mu_{0j}\left(t,x\right)\right)\right\}^2]E\left[\left(\Delta_2\left(t_2\right)-\Lambda_2\left(X_2\left(t_2\right)\right)\right)^2\right] = \frac{v_0\nu_2}{\nu_0+\nu_2}\left(1-e^{-\lambda_2(t_2-Y)}\right).
$$

Summary:

1) Covariance of log-rank test statistics in the case of bivariate exponential distribution is

$$\Omega = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix},$$

where

$$\begin{aligned} \sigma_{11} &= \frac{v_1 \nu_2}{\nu_1 + \nu_2} \left(1 - e^{-\lambda_1 (t_1 - Y)}\right) \\ \sigma_{12} &= \frac{v_0 \nu_1 \nu_2}{(\nu_0 + \nu_1)(\nu_1 + \nu_2)} \left(\frac{\lambda_2}{\lambda_1} \left(1 - e^{-\lambda_1 (t_1 - Y)}\right)\right) \\ \sigma_{22} &= \frac{v_0 \nu_2}{\nu_0 + \nu_2} \left(1 - e^{-\lambda_2 (t_2 - Y)}\right), \text{ for } 0 \leqslant t_1 \leqslant t_2. \end{aligned}$$

2) If assume that patients have the same entry time ($Y = 0$), then

$$\begin{aligned} \sigma_{11} &= \frac{v_1 \nu_2}{\nu_1 + \nu_2} \left(1 - e^{-\lambda_1 t_1}\right) \\ \sigma_{12} &= \frac{v_0 \nu_1 \nu_2}{(\nu_0 + \nu_1)(\nu_1 + \nu_2)} \left(\frac{\lambda_2}{\lambda_1} \left(1 - e^{-\lambda_1 t_1}\right)\right) \\ \sigma_{22} &= \frac{v_0 \nu_2}{\nu_0 + \nu_2} \left(1 - e^{-\lambda_2 t_2}\right), \text{ for } 0 \leqslant t_1 \leqslant t_2. \end{aligned}$$

3) If assume that patient entry time $Y \sim h(y)$, then

$$\begin{aligned} \sigma_{11} &= \int_0^\infty \frac{v_1 \nu_2}{\nu_1 + \nu_2} \left(1 - e^{-\lambda_1 (t_1 - y)}\right) h(y)\, dy \\ &= \frac{v_1 \nu_2}{\nu_1 + \nu_2} \left(1 - \int_0^\infty e^{-\lambda_1 (t_1 - y)} h(y)\, dy\right) \\ \sigma_{12} &= \int_0^\infty \frac{v_0 \nu_1 \nu_2}{(\nu_0 + \nu_1)(\nu_1 + \nu_2)} \left(\frac{\lambda_2}{\lambda_1} \left(1 - e^{-\lambda_1 (t_1 - y)}\right)\right) h(y)\, dy \\ &= \frac{v_0 \nu_1 \nu_2}{(\nu_0 + \nu_1)(\nu_1 + \nu_2)} \frac{\lambda_2}{\lambda_1} \left(1 - \int_0^\infty e^{-\lambda_1 (t_1 - y)} h(y)\, dy\right) \\ \sigma_{22} &= \int_0^\infty \frac{v_0 \nu_2}{\nu_0 + \nu_2} \left(1 - e^{-\lambda_2 (t_2 - y)}\right) h(y)\, dy \\ &= \frac{v_0 \nu_2}{\nu_0 + \nu_2} \left(1 - \int_0^\infty e^{-\lambda_2 (t_2 - y)} h(y)\, dy\right), \text{ for } 0 \leqslant t_1 \leqslant t_2. \end{aligned}$$

Chapter 11

Two-stage winner designs with survival outcomes

In this section, two-stage winner design with survival outcomes using the same or different survival endpoints at the interim and final analysis is studied.

11.1 Notations

The following notations will be used in this chapter:

$\nu_0 : \nu_1 : \nu_2$: allocation ratio for control arm, treatment arm 1, and treatment arm 2,

where $\nu_0 + \nu_1 + \nu_2 = 1$

ν : for simplicity, set $\nu_1 = \nu_2 = \upsilon$

ω_1 : hazard ratio for treatment 1 versus control

ω_2 : hazard ratio for treatment 2 versus control

N_{1i} : number of patients at risk at the 1st stage in treatment 1 when i^{th} event occurred

N_{2i} : number of patients at risk at the 1st stage in treatment 2 when i^{th} event occurred

$N_{0i}^{(2)}$: number of patients at risk at the 2^{nd} stage in control arm when i^{th} event occurred

$N_{ji}^{(2)}$: number of patients at risk at the 2^{nd} stage in the winning treatment arm j when i^{th} event occurred

d_0 : total number of events in the two treatment arms in the first stage

d_1 : total number of events in the control arm and winning treatment arm at the end of the first stage

d_2 : total number of events in the control arm and winning treatment arm at the end of the second stage

τ : information time for stage 1, $\tau = d_1/d_2$

11.1.1 Log-rank test statistics and its distribution

Assuming no ties, the standardized log-rank test statistic to compare the two treatments at the end of stage 1 is

$$LR_1 = \frac{\sum_{i=1}^{d_0}(I_{2i} - \frac{N_{2i}}{N_{1i}+N_{2i}})}{\sqrt{\sum_{i=1}^{d_0}\frac{N_{1i}N_{2i}}{(N_{1i}+N_{2i})^2}}},$$

where $I_{2i} = 1$ if the event occurred in treatment arm 2 and 0 otherwise. For fixed d_0, when assuming $\omega_1 \neq \omega_2$, LR_1 is approximately normally distributed with unit variance and mean (Wassmer 2006)

$$\lambda \equiv E(LR_1) = \sqrt{d_0}\frac{\sqrt{\nu_1\nu_2}}{\nu_1+\nu_2}\ln(\omega_2/\omega_1).$$

The log-rank test statistic comparing control arm and winning arm at the end of stage 2 is

$$LR_{0j}^{(2)} = \frac{\sum_{i=1}^{d_2}(I_{0i} - \frac{N_{0i}^{(2)}}{N_{ji}^{(2)}+N_{0i}^{(2)}})}{\sqrt{\sum_{i=1}^{d_2}\frac{N_{ji}^{(2)}N_{0i}^{(2)}}{(N_{ji}^{(2)}+N_{0i}^{(2)})^2}}},$$

where $I_{0i} = 1$ if the event occurred in the control arm and 0 otherwise. For fixed d_1, d_2, $LR_{0j}^{(1)}$ and $LR_{0j}^{(2)}$ are approximately normally distributed with unit variance and means

$$E(LR_{0j}^{(2)}) = \sqrt{d_2}\frac{\sqrt{\nu_0\nu_j}}{\nu_0+\nu_j}\ln(1/\omega_j).$$

At the interim look, the following decision rules will be used to pick the winning arm: drop treatment 1, when $LR_1 < 0$; drop treatment 2 when $LR_1 > 0$. If we denote the interim statistic as

$$V = LR_1,$$

then the final test statistic of the two-stage winner design will be

$$W = \begin{cases} LR_{01}^{(2)} & \text{if } V \geq 0, \\ LR_{02}^{(2)} & \text{if } V < 0. \end{cases}$$

The winning probability of treatment 1 can be calculated as in Shun et al (2008):

$$p = P(V \geq 0) \doteq P(V - \lambda > -\lambda) = 1 - \Phi(-\lambda) = \Phi(\lambda),$$

where $\Phi(\cdot)$ is the standard normal cdf. Note that both V and $LR_{0j}^{(1)}$ are asymptotically normally distributed with asymptotically unit variance. Since they both summarize information of two arms and have one common arm, they are correlated.

11.2 Approximate Parameter Calculations

In practice, when a clinical trial is designed, one of the considerations is to chose the appropriate winning probability (p) and the time doing the interim analysis $(\tau = d_1/d_2)$. The inferior treatment group is terminated at the interim stage to increase the efficiency (saving time and resources) of the clinical trial. The earlier the interim analysis is, the more efficiency the design may have. On the other hand, the later the interim analysis is, the more reliable the interim treatment selection is. In the section below, some practical considerations are provided for the approximate calculation of sample sizes and critical values.

11.2.1 Fix the winning probability approach

A) Total number of events d_0 in the two treatment arms at interim analysis:

Since

$$P\left(V \geq 0\right) \doteq \Phi\left(\lambda\right) = \Phi\left(-\sqrt{d_0}\frac{\sqrt{\nu_1\nu_2}}{\nu_1+\nu_2}\ln(\omega_1/\omega_2)\right) = p,$$

we have

$$d_0 = \frac{\left(\nu_1+\nu_2\right)^2}{\nu_1\nu_2}\left(z_p/\ln(\omega_1/\omega_2)\right)^2,$$

where p is the winning probability.

B) Interim time t_1 and total number of events d_1 in the winning arm and control arm at interim analysis:

The total event number of events d_1 in the winning treatment arm and control arm is determined by the time t_1 when d_0 events are observed in the two treatment arms. In order to calculate d_1, one needs to know when the expected total event number of the two treatment groups reaches d_0.

Denote the cdf for the event time of the control arm and the two treatment arms as $F_0(\cdot)$, $F_1(\cdot)$ and $F_2(\cdot)$, respectively. Assume a constant accrual rate of one patient per a unit time. Then at t_1, the expected total number of events in the two treatment arms can also be denoted as

$$d_0 = \sum_{i \le t_1/a} [\nu_1 F_1(t_1 - ai) + \nu_2 F_2(t_1 - ai)].$$

Since d_0 is already determined, by solving this equation, we can get the t_1.

C) Total number of events in the winning arm and control arm at anytime $t_2 \ge t_1$:

The expected event number for the winning arm and the control arm at time $t_2 \ge t_1$ is approximately calculated as

$$d_1 = \sum_{i \le t_1/a} [p\nu_1 F_1(t_1 - ai) + q\nu_2 F_2(t_1 - ai) + \nu_0 F_0(t_1 - ai)],$$

$$\begin{aligned} d_2 = & \sum_{i \le t_1/a} [p\nu_1 F_1(t_2 - ai) + q\nu_2 F_2(t_2 - ai) + \nu_0 F_0(t_2 - ai)] \\ & + \sum_{i \le (t_2 - t_1)/a} [\frac{p\nu}{\nu + \nu_0} F_1(t_2 - t_1 - ai) + \frac{q\nu}{\nu + \nu_0} F_2(t_2 - t_1 - ai) + \frac{\nu_0}{\nu + \nu_0} F_0(t_2 - t_1 - ai)] \end{aligned}$$

where $q = 1 - p$.

D) Wasted number of events and number of patients in the dropped arm:

The wasted number of events in the dropped arm at the interim analysis is

$$\sum_{i \le t_1/a} [q\nu_1 F_1(t_1 - ai) + p\nu_2 F_2(t_1 - ai)].$$

And the wasted number of patients in the dropped arm is

$$n_0 = \nu t_1 / a.$$

E) Critical value and power:

For a given type I error rate, the critical value and power of the design can be calculated by using the bivariate normal distribution function. Note that the critical value may depend on the information time $\tau = d_1/d_2$. However, the power will generally increase as d_2 increases, therefore, a unique d_2 value can be found for a given power and the type I error requirement.

F) Examples

I) Same endpoint at interim and final analysis In this example, overall survival (OS) is the endpoint of interest. An accrual rate of 23 patients per month is used for the calculation. The allocation ratio is $1:1:1$ in the first stage and 1:1 in the second stage. Assume the median survival for OS of the control group is 7.5 months. The hazard ratio for treatment 1 versus control is fixed at $\omega_1 = 0.8$. The hazard ratio ω_2 for treatment 2 versus control used in the calculations varies from 0.5 to 0.75 with an increment of 0.05. The desired one-sided type I error and power used in the calculation is 0.025 and 0.9 respectively. Based on these assumptions, the following parameters are calculated and presented in Table 11.1

and Figures 11.1 to 11.7: Information time τ, total number of events in the two treatment arms d_0, total number of events in the control and winning arm at stage 1 d_1 and at stage 2 d_2, critical value at final analysis, and the time (months) required to finished the first stage and second stage.

As in Table 11.1, when the winning probability is 0.7 , d_0 is only 5 when $\omega_2 = 0.50$. When ω_2 increases to 0.70 with increment of 0.05, d_0 increases slowly to 61.7. When ω_2 reach 0.75, d_0 jumped to 264.1. The same trend is observed when the winning probability is 0.8 and 0.85. For a fixed winning probability, when the treatment difference between the two treatment arms is large, the required number of events to pick the winner arm at the interim analysis will be very small. When the treatment effects of two treatment arms are very close, the number of events required to pick a winner increases dramatically. Information time τ, d_1, critical value, interim analysis time are all positive correlated to d_0. When ω_2 is 0.75, to maintain a winning probability of 0.8 at the interim analysis, d_0 is 1.4213 times of the total number of events required at the final analysis. This indicates that when the winning probability is very large, the required number of events could be larger than the total number of events required at the end of the study. In this case, two-stage winner design will not be applicable any more. Thus, there exists a maximum winning probability for pre-specified ω_1 and ω_2. This maximum winning probability will be reached when the information time at the interim is 1. The maximum winning probabilities are presented in Table 11.2 and Figure 11.8 for several scenarios.

Table 11.1: Two-stage Winner Designs Using Same Endpoints When $\omega_1 = 0.8$

ω_2	Parameter	Winner Probability		
		p=0.7	p=0.8	p=0.85
0.5	τ	0.0133	0.0526	0.1300
	d_0	5.0	12.8	19.5
	d_1	6.0	15.0	22.4
	d_2	450.2	285.0	172.3
	critical value	2.0039	2.0434	2.0841
	interim time(months)	3.5	5.7	7.0
	final time(months)	33.6	26.0	19.9
0.55	τ	0.0206	0.0823	0.1757
	d_0	7.8	20.2	30.6
	d_1	9.2	23.2	34.7
	d_2	447.7	282.0	197.6
	critical value	2.0149	2.0618	2.1006
	interim time(months)	4.3	7.0	8.8
	final time(months)	33.4	26.0	21.8
0.6	τ	0.0346	0.1311	0.2436
	d_0	13.3	34.2	51.9
	d_1	15.4	38.7	57.9
	d_2	444.6	295.0	237.7
	critical value	2.0289	2.0845	2.1201
	interim time(months)	5.6	9.2	11.6
	final time(months)	33.3	27.0	24.6
0.65	τ	0.0645	0.2162	0.3677
	d_0	25.5	65.7	99.7
	d_1	28.9	72.7	108.9
	d_2	448.8	336.3	296.2
	critical value	2.0514	2.1128	2.1468
	interim time(months)	7.7	13.0	16.6
	final time(months)	34.0	30.0	28.9
0.7	τ	0.1401	0.4186	0.6740
	d_0	61.7	158.9	241.0
	d_1	68.2	171.2	256.2
	d_2	486.9	409.0	380.1
	critical value	2.0880	2.1555	2.1876
	interim time(months)	12.4	21.6	28.2
	final time(months)	37.0	35.9	35.9
0.75*	τ	0.4928	**1.4213**	
	d_0	264.1	680.2	
	d_1	279.8	701.2	
	d_2	567.8		
	critical value	2.1666		
	interim time(months)	29.6		
	final time(months)	45.6		

* When p=0.8 or 0.85, number of events required at interim is larger than the final analysis.

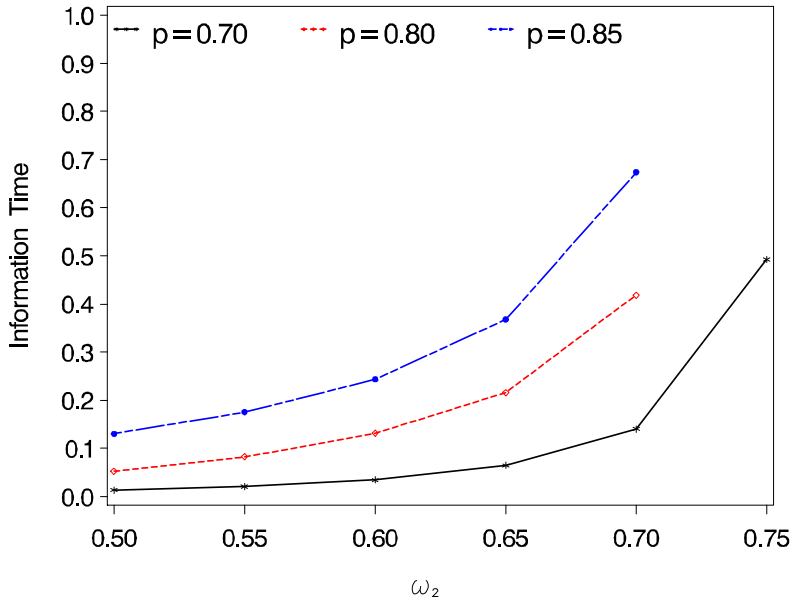

Figure 11.1: Information Time for Two-stage Winner Design with Same Endpoint when $\omega_1 = 0.8$

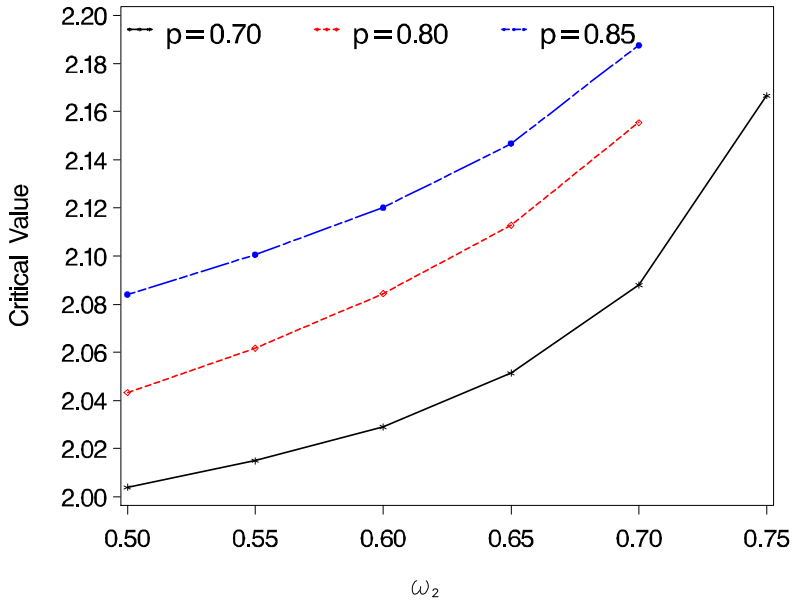

Figure 11.2: Critical Values for Two-stage Winner Designs with Same Endpoint when $\omega_1 = 0.8$

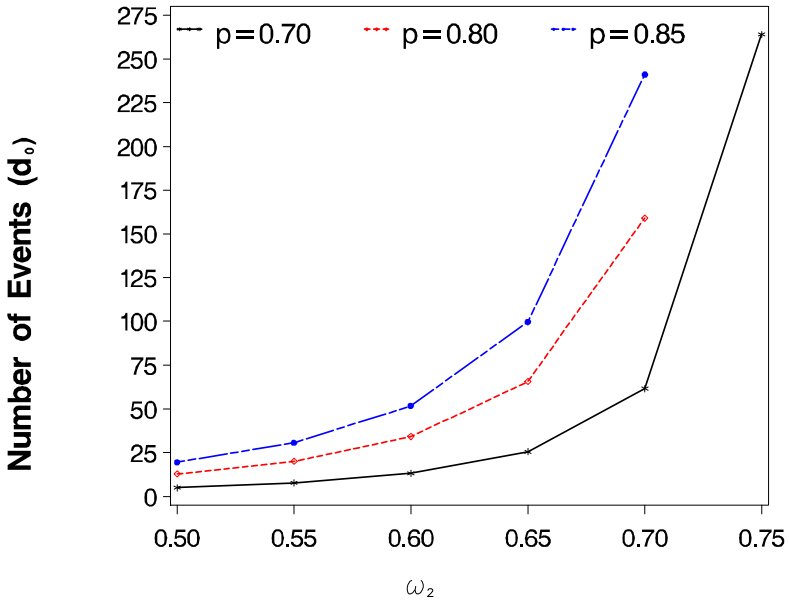

Figure 11.3: Total Events in Treatment Groups for Two-stage Winner Design with Same Endpoint when $\omega_1 = 0.8$

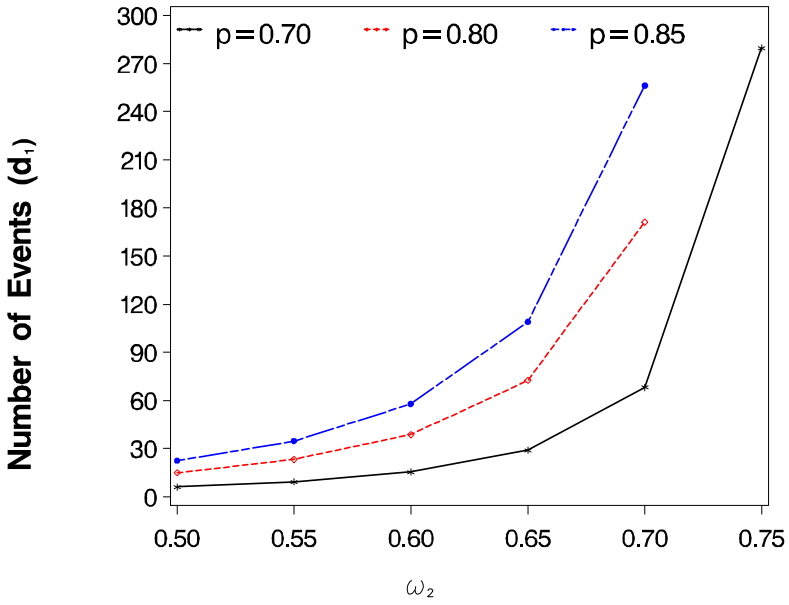

Figure 11.4: Total Events in Winning and Control Groups at First Stage for Two-stage Winner Designs with Same Endpoint when $\omega_1 = 0.8$

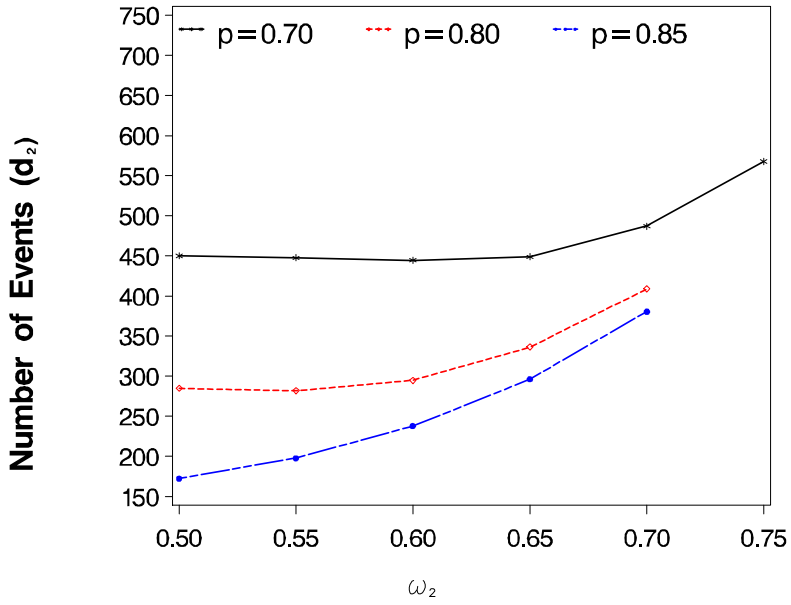

Figure 11.5: Total Events in winning and Control Groups at Final Stage for Two-stage Winner Designs with Same Endpoint when $\omega_1 = 0.8$

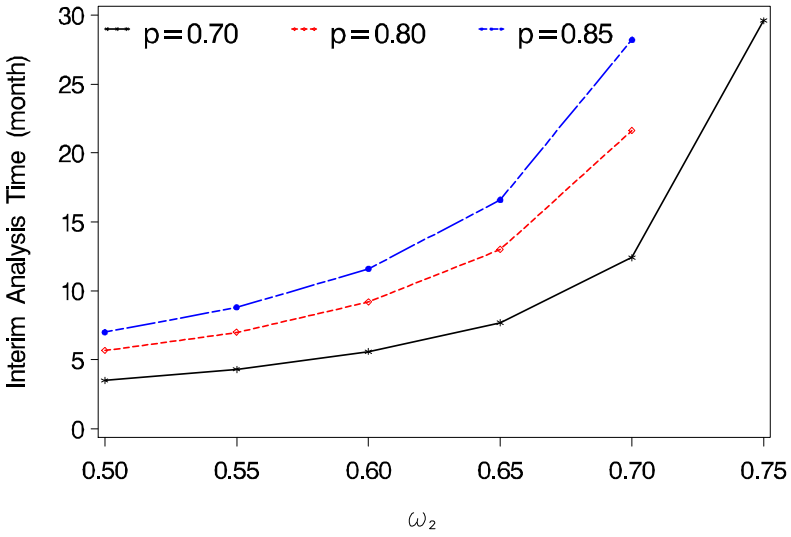

Figure 11.6: Interim Analysis Time for Two-stage Winner Designs with Same Endpoint when $\omega_1 = 0.8$

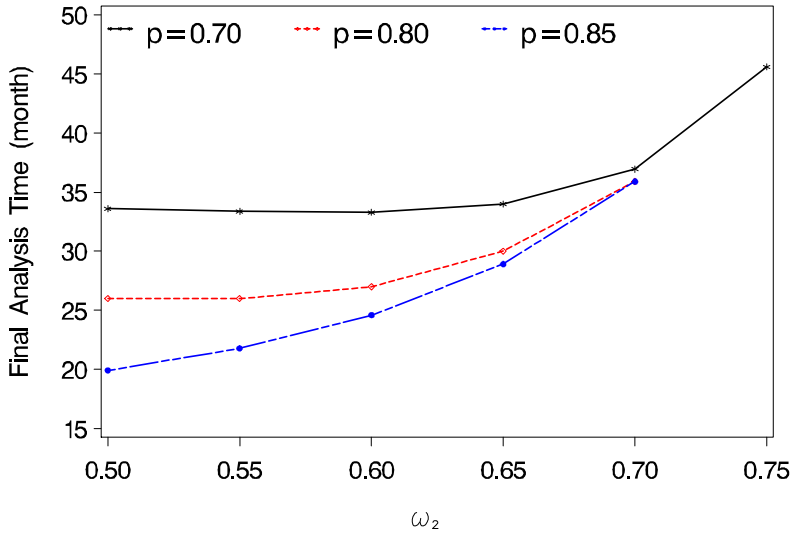

Figure 11.7: Final Analysis Time for Two-stage Winner Designs with Same Endpoint when $\omega_1 = 0.8$

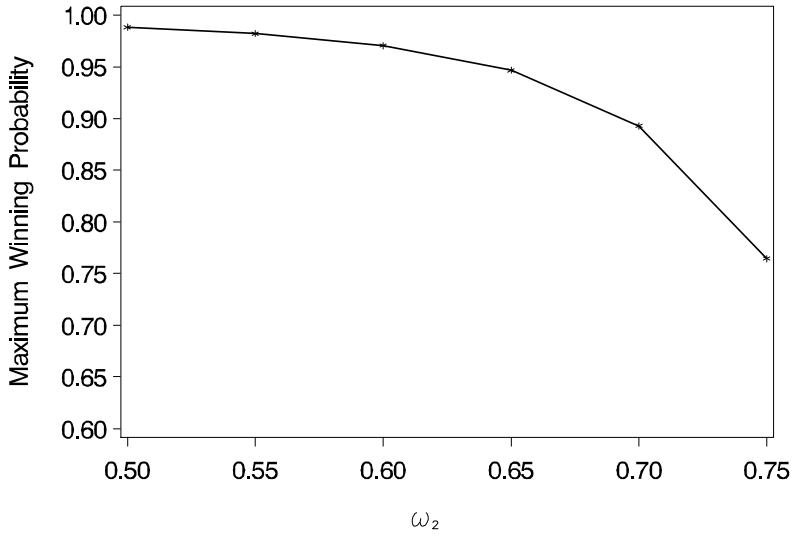

Figure 11.8: Maximum Winning Probability for Two-stage Winner Designs with Same Endpoint when $\omega_1 = 0.8$

Table 11.2: Two-stage Winner Designs with Maximum Winning Probability when $\omega_1 = 0.8$

	ω_2					
Parameter	**0.50**	**0.55**	**0.60**	**0.65**	**0.70**	**0.75**
maximum winning probability	0.9882	0.9821	0.9705	0.9467	0.8928	0.7645
d_0	92.8	125.4	172.4	241.7	345.8	498.9
d_1	101.2	136.6	184.7	256.5	363.3	518.7
d_2	101.2	135.6	184.7	256.5	363.4	518.7
critical value	2.2121	2.2121	2.2121	2.2121	2.2121	2.2121
interim time(months)	16.7	19.6	23.4	28.6	35.9	46.1
final time(months)	16.7	19.6	23.4	28.6	35.9	46.1

Two-stage winner design property based on various winning probability was also explored (see Table 11.3). The hazard ratios used for the calculation are $\omega_1 = 0.8$ and $\omega_2 = 0.70$. The winning probability cannot be smaller than 0.5. So the the winning probability used for calculations starts from 0.5. In Table 11.3, when $p = 0.5$, d_0 is zero. When p is getting larger, d_0, information time, critical values, and interim time at final analysis are also getting larger until p reaches its limit. However the final analysis time will not be shorten any more when winning probability is larger than 0.7.

Table 11.3: Two-stage Winner Designs with Different Winning Probabilities when $\omega_1 = 0.8$ and $\omega_2 = 0.7$

	Winning Probability					
Parameter	p=0.6	p=0.7	p=0.8	p=0.85	p=0.8927923	p=0.9
τ	0.0286	0.1401	0.4186	0.6740	1	1.0691*
d_0	14.4	61.7	158.9	241.0	345.8	
d_1	16.3	68.2	171.2	256.2	363.3	
d_2	570.9	486.9	409.0	380.1	363.4	
critical value	2.0230	2.0880	2.1555	2.1876	2.2121	
interim time(months)	5.6	12.4	21.6	28.2	35.9	
final time(months)	38.7	37.0	35.9	35.9	35.9	

Number of events required at interim analysis is larger than at final analysis

To simplify the calculation, the results shown above is based on the assumption

that the accrual will not stop until the target event number is reached. When accrual is stopped earlier, longer time will be needed to observe the same number of events as shown in Table 11.4. The earlier the accrual is stopped, the longer time is needed to finish the study.

Table 11.4: Two-stage Winner Designs with Different Accrual Stopping Time when $\omega_1 = 0.8$ and $\omega_2 = 0.7$

	p=0.7		p=0.8		p=0.85	
Accrual stopped at percentage of total events	accrual time (months)	final time (months)	accrual time (months)	final time (months)	accrual time (months)	final time (months)
0.7	29.8	41.3	29.3	40.0	28.9	42.0
0.8	32.2	38.4	31.6	37.2	31.4	37.6
0.9	34.6	37.3	33.8	36.2	33.7	36.3
1.0	37.0	37.0	35.9	35.9	35.9	35.9

The effect of different accrual rates on the two-stage winner design was also studied. As shown in Table 11.5, when the accrual rate was as low as 10 patients per month, both the interim time and final time had to be sufficiently long to reach the target event number. On the contrary, when a study has a fast enrollment of 50 patients per month, both the interim and final time is shortened. When winning probability is 0.7, it only takes 8 months to do the interim analysis. When the difference between two hazard ratio is larger, this time will be further shortened. However, when the target winning probability is high (0.85), the interim time will take 17.4 months which is only 4 months to the end of 2nd stage. This indicates that when accrual rate is high the two-stage winner design may not be a good choice when a high winning probability is required.

Table 11.5: Two-stage Winner Designs with Different Accrual Rates When $\omega_1 = 0.8$ and $\omega_2 = 0.7$

Accrual Rate	Parameter	Winner Probability		
		p=0.7	p=0.8	p=0.85
10 per month	τ	0.1374	0.4080	0.6565
	d_0	61.7	158.9	241.0
	d_1	66.9	167.1	250.0
	d_2	487.2	409.6	380.9
	critical value	2.0870	2.1538	2.1859
	interim time(months)	20.3	37.3	50.3
	final time(months)	68.2	65.8	66.6
23 per month	τ	0.1401	0.4186	0.6740
	d_0	61.7	158.9	241.0
	d_1	68.2	171.2	256.2
	d_2	486.9	409.0	380.1
	critical value	2.0880	2.1555	2.1876
	interim time(months)	12.4	21.6	28.2
	final time(months)	37.0	35.9	35.9
50 per month	τ	0.1420	0.4268	0.6889
	d_0	61.7	158.9	241.0
	d_1	69.1	174.3	261.4
	d_2	486.7	408.5	379.5
	critical value	2.0887	2.1568	2.1891
	interim time(months)	8.0	13.6	17.4
	final time(months)	22.2	21.6	21.5

II) Different endpoints Fix the winning probability approach was used to do the calculation when surrogate endpoint was used at the interim analysis. The formulas used when the same endpoint used at interim and final are still applicable here. Calculations were done for the following scenarios: 1) fix the two hazard ratios for OS and one of the hazard ratios for PFS and let the other hazard ratio for PFS vary from 0.5 to 0.75 by an increment of 0.05 (See Table 11.6 and Figures 11.9 to 11.15). 2) fix the two hazard ratios for PFS and one of the hazard ratios for OS and let the other hazard ratio for OS vary from 0.5 to 0.75 by an increment of 0.05 (See Table 11.7 and Figures 11.16 to 11.22). The similar trends were

observed for relationships between the different design parameters. When the magnitude of PFS difference between two treatment group is larger than the difference of OS, the total number of PFS events required at the interim analysis was very small as shown in Table 11.6. The larger the PFS difference was, the less the PFS events was required. However, when the difference of PFS was smaller than the difference of OS, the number of PFS events at the interim analysis was increased substantially. When both the hazard ratios for PFS were fixed, the expected number of PFS events required at the interim analysis were fixed (See Table 11.7).

Table 11.6: Two-stage Winner Designs Using Different Endpoints When OS $\omega_1 = 0.8$ and $\omega_2 = 0.7$ and PFS $\omega_2 = 0.8$

PFS ω_1	Parameter	Winner Probability		
		p=0.7	p=0.8	p=0.85
0.50	τ	0.0037	0.0114	0.0191
	d_0	5.0	12.9	19.5
	d_1	1.9	5.0	7.8
	d_2	508.8	439.6	408.8
	critical value	1.9715	1.9793	1.9841
	interim time(months)	1.9	3.1	3.8
	final time(months)	34.6	31.8	30.6
0.55	τ	0.0056	0.0180	0.0307
	d_0	7.8	20.2	30.6
	d_1	2.9	7.9	12.5
	d_2	508.3	439.5	408.6
	critical value	1.9740	1.9834	1.9893
	interim time(months)	2.3	3.9	4.9
	final time(months)	34.7	32.0	30.9
0.60	τ	0.0096	0.0320	0.0546
	d_0	13.3	34.6	51.9
	d_1	4.9	14.0	22.3
	d_2	507.9	439.2	408.6
	critical value	1.9777	1.9897	1.9967
	interim time(months)	3.0	5.2	6.7
	final time(months)	35.0	32.5	31.5
0.65	τ	0.0191	0.0656	0.1162
	d_0	25.5	65.7	99.7
	d_1	9.7	28.8	47.4
	d_2	507.1	438.4	408.1
	critical value	1.9838	1.9992	2.0086
	interim time(months)	4.3	7.6	10.1
	final time(months)	35.3	33.2	32.5
0.70	τ	0.0517	0.1925	0.3537
	d_0	61.7	158.9	241.0
	d_1	26.2	84.1	144.1
	d_2	506.0	436.8	407.4
	critical value	1.9954	2.0176	2.0315
	interim time(months)	7.2	14.0	19.4
	final time(months)	36.2	35.1	35.2
0.75	τ	0.3227	0.9999	0.9999
	d_0	264.1	680.2	1031.6
	d_1	161.7	542.0	888.0
	d_2	501.0	542.0	888.0
	critical value	2.0277	2.0594	2.0565
	interim time(months)	20.8	48.0	70.9
	final time(months)	40.2	48.0	70.9

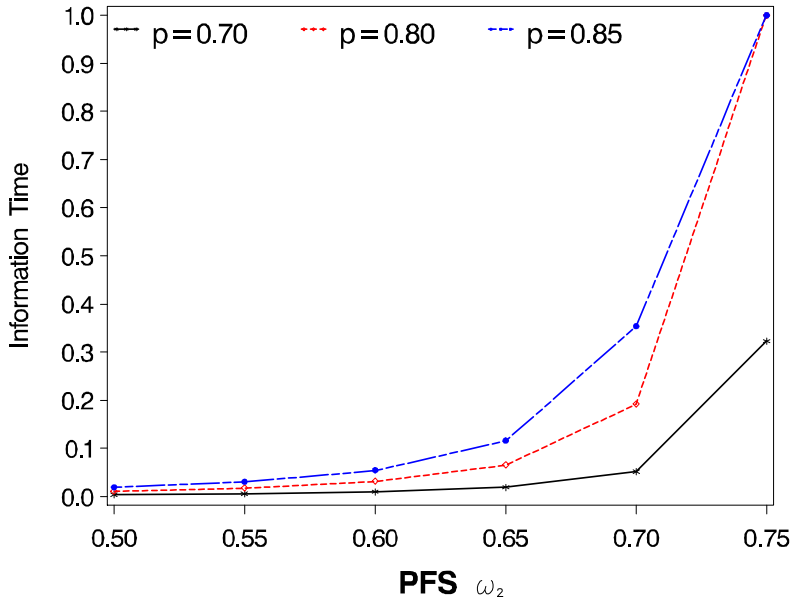

Figure 11.9: Information Time for two-stage Winner Designs with Different Endpoints when OS ω_1=0.8 and $\omega_2 = 0.7$ and PFS $\omega_1 = 0.8$

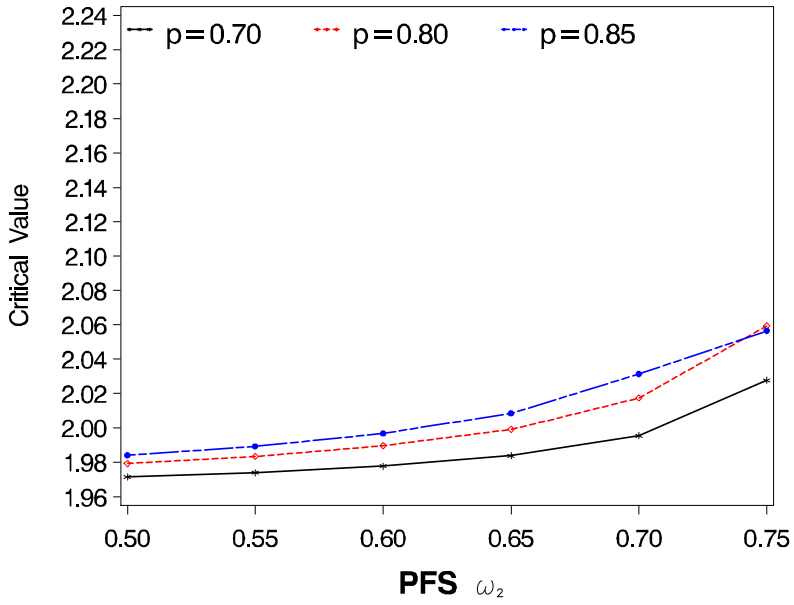

Figure 11.10: Critical Values for two-stage Winner Designs with Different Endpoints when OS ω_1=0.8 and $\omega_2 = 0.7$ and PFS $\omega_1 = 0.8$

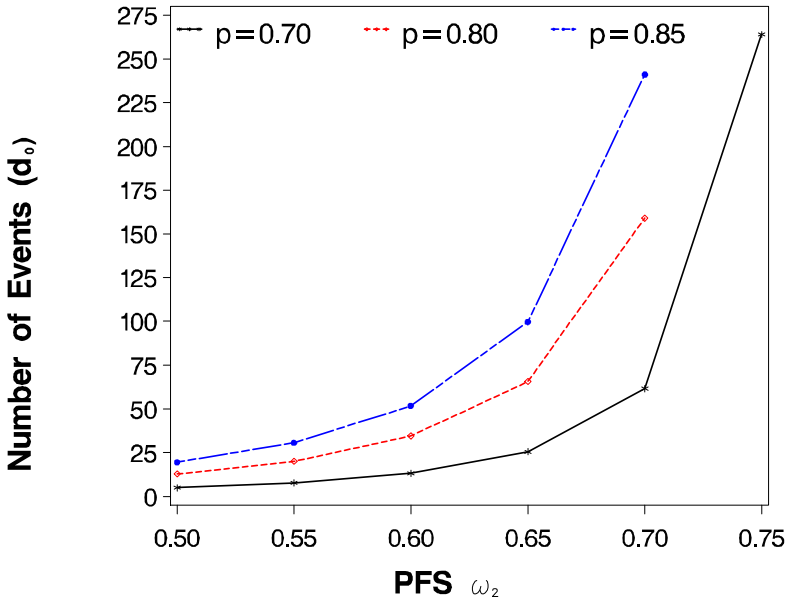

Figure 11.11: Total Events in Two Treatment Groups for Two-stage Winner Designs with Different Endpoints when OS ω_1=0.8 and $\omega_2 = 0.7$ and PFS $\omega_1 = 0.8$

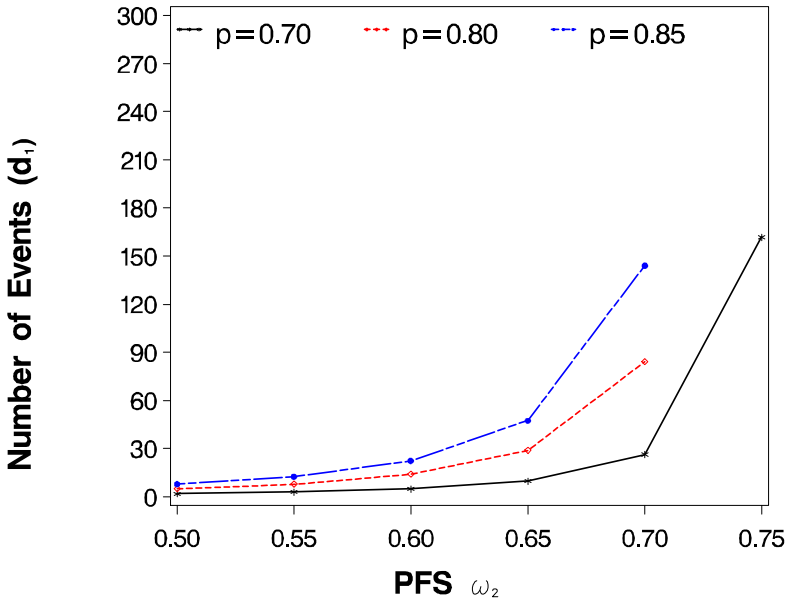

Figure 11.12: Total Events in Winner and Control Groups at First Stage for Two-stage Winner Designs with Different Endpoints when OS ω_1=0.8 and $\omega_2 = 0.7$ and PFS $\omega_1 = 0.8$

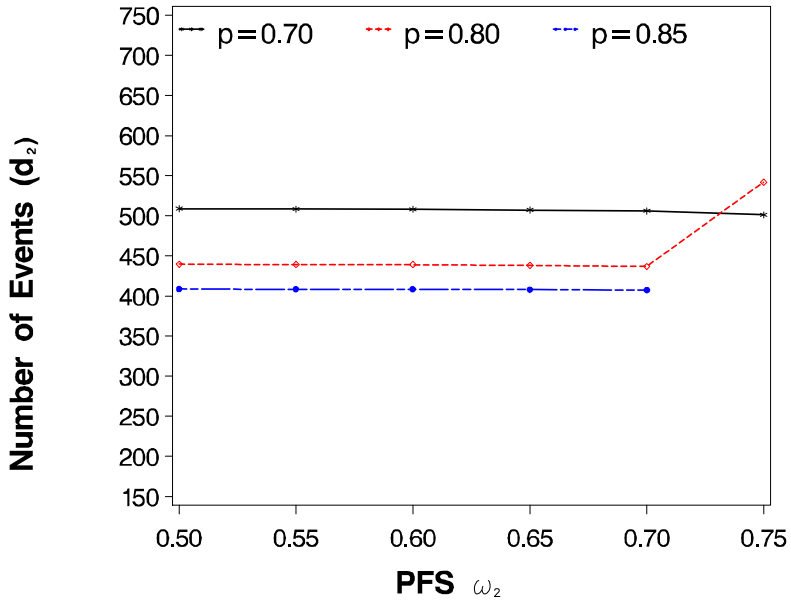

Figure 11.13: Total Events in Winner and Control Groups at Final Stage for Two-stage Winner Designs with Different Endpoints when OS ω_1=0.8 and $\omega_2 = 0.7$ and PFS $\omega_1 = 0.8$

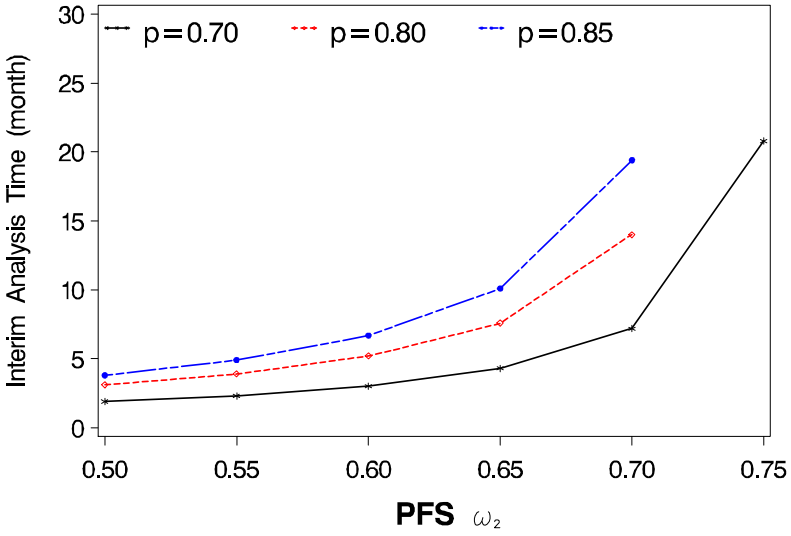

Figure 11.14: Interim Analysis Time for Two-stage Winner Designs with Different Endpoints when OS ω_1=0.8 and $\omega_2 = 0.7$ and PFS $\omega_1 = 0.8$

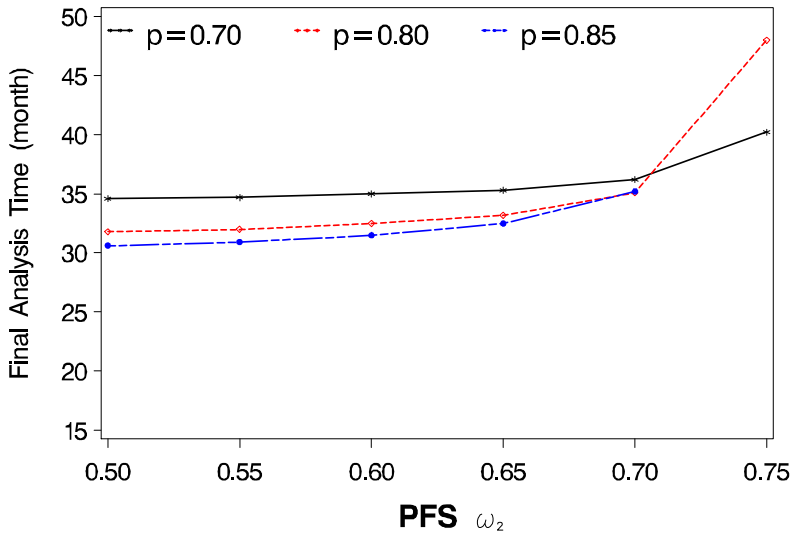

Figure 11.15: Final Analysis Time for Two-stage Winner Designs with Different Endpoints when OS ω_1=0.8 and $\omega_2 = 0.7$ and PFS $\omega_1 = 0.8$

Table 11.7: Two-stage Winner Designs Using Different Endpoints When PFS $\omega_1 = 0.8$ and $\omega_2 = 0.7$ and OS $\omega_1 = 0.8$

OS ω_1	Parameter	Winner Probability		
		p=0.7	p=0.8	p=0.85
0.50	τ	0.0536	0.2665	0.7494
	d_0	61.7	158.9	241.0
	d_1	24.2	77.4	132.6
	d_2	452.0	290.3	177.0
	critical value	1.9963	2.0271	2.0597
	interim time(months)	7.2	14.0	19.4
	final time(months)	34.9	28.8	23.0
0.55	τ	0.0548	0.2706	0.6629
	d_0	61.7	158.9	241.0
	d_1	24.7	79.2	135.7
	d_2	451.5	292.5	204.7
	critical value	1.9966	2.0275	2.0545
	interim time(months)	7.2	14.0	19.4
	final time(months)	34.6	28.6	24.6
0.6	τ	0.0558	0.2613	0.5592
	d_0	61.7	158.9	241.0
	d_1	25.2	80.9	138.7
	d_2	452.4	309.5	247.9
	critical value	1.9968	2.0264	2.0477
	interim time(months)	7.2	14.0	19.4
	final time(months)	34.3	29.2	27.1
0.65	τ	0.0557	0.2328	0.4544
	d_0	61.7	158.9	241.0
	d_1	25.7	82.5	141.4
	d_2	461.4	354.3	311.3
	critical value	1.9967	2.0229	2.04
	interim time(months)	7.2	14.0	19.4
	final time(months)	34.4	31.3	30.5
0.70	τ	0.0517	0.1925	0.3537
	d_0	61.7	158.9	241.0
	d_1	26.2	84.1	144.1
	d_2	506.0	436.8	407.4
	critical value	1.9954	2.0176	2.0315
	interim time(months)	7.2	14.0	19.4
	final time(months)	36.2	35.1	35.2
0.75	τ	0.0433	0.1477	0.2600
	d_0	61.7	158.9	241.0
	d_1	26.6	85.6	146.6
	d_2	615.4	579.5	563.7
	critical value	1.9924	2.0109	2.0220
	interim time(months)	7.2	14.0	19.4
	final time(months)	41.1	41.6	42.5

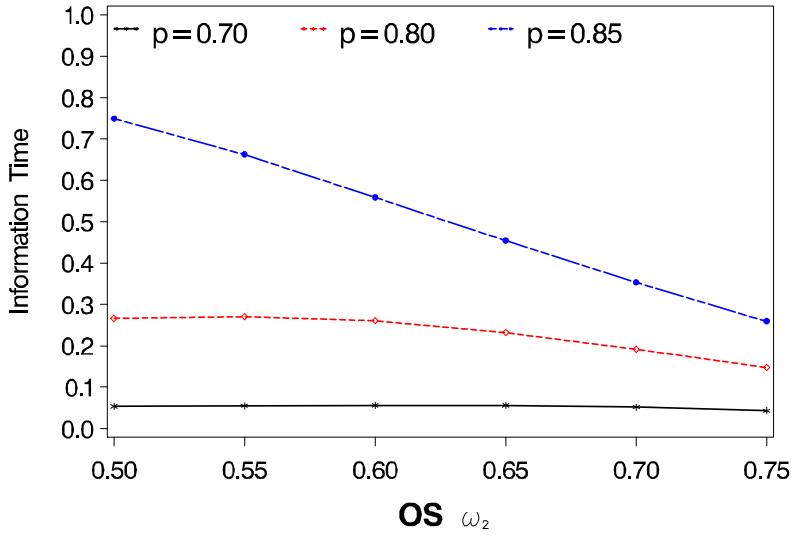

Figure 11.16: Information Time for Two-stage Winner Designs with Different Endpoints when PFS ω_1=0.8 and $\omega_2 = 0.7$ and OS $\omega_1 = 0.8$

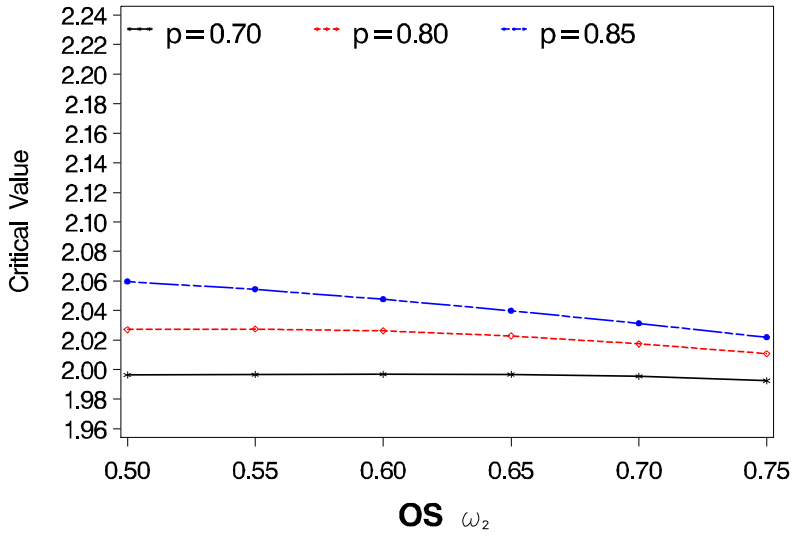

Figure 11.17: Critical Values for Two-stage Winner Designs with Different Endpoints when PFS ω_1=0.8 and $\omega_2 = 0.7$ and OS $\omega_1 = 0.8$

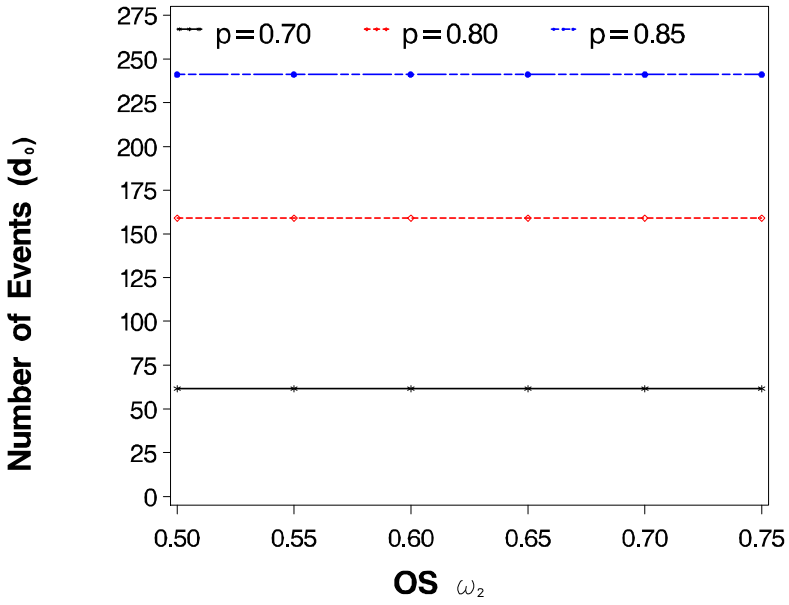

Figure 11.18: Total Events in Two Treatment Groups for Two-stage Winner Designs with Different Endpoints when PFS ω_1=0.8 and $\omega_2 = 0.7$ and OS $\omega_1 = 0.8$

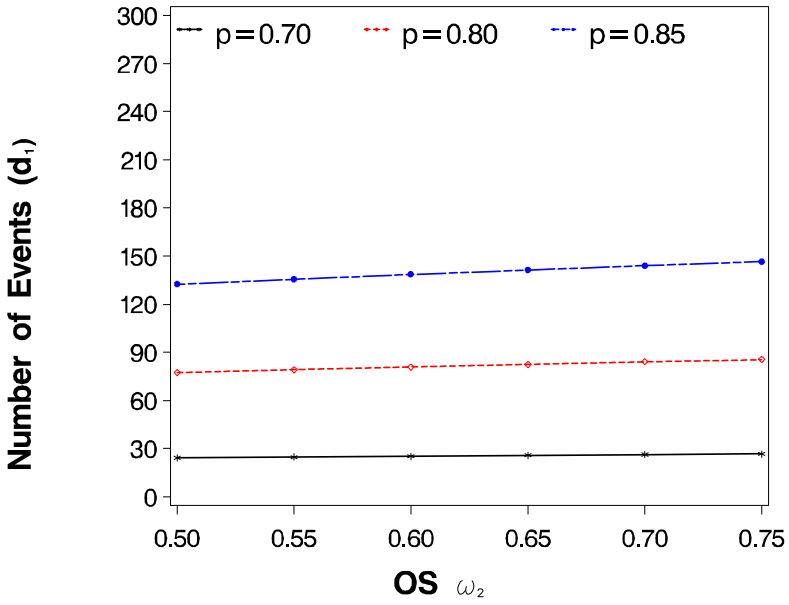

Figure 11.19: Total Events in Winner and Control Groups at First Stage for Two-stage Winner Designs with Different Endpoints when PFS ω_1=0.8 and $\omega_2 = 0.7$ and OS $\omega_1 = 0.8$

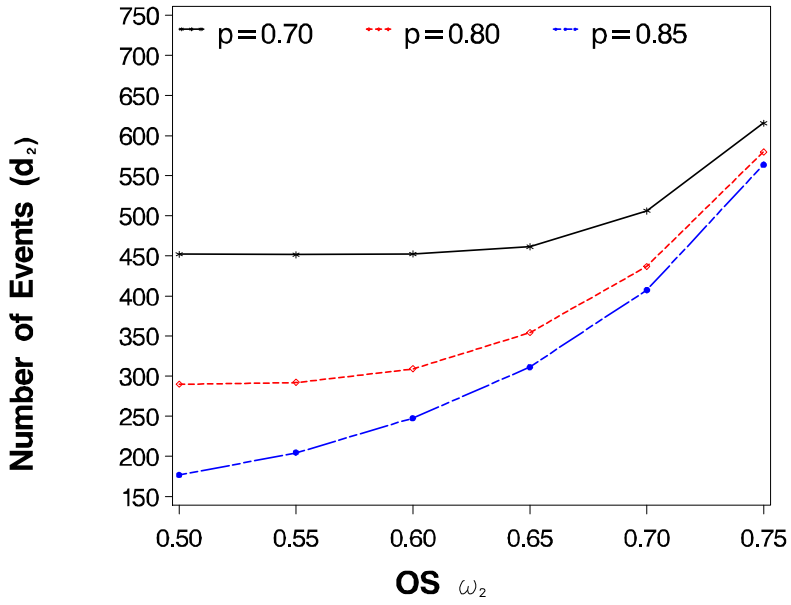

Figure 11.20: Total Events in Winner and Control Groups at Final Stage for Two-stage Winner Designs with Different Endpoints when PFS ω_1=0.8 and $\omega_2 = 0.7$ and OS $\omega_1 = 0.8$

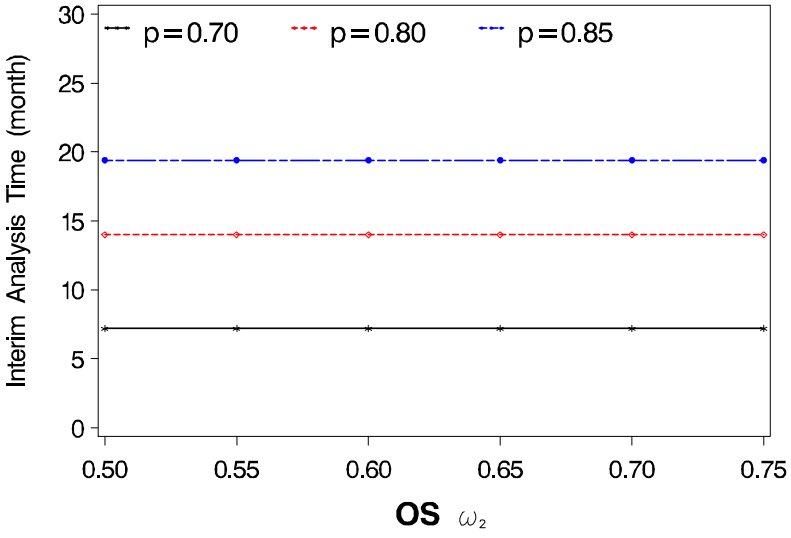

Figure 11.21: Interim Analysis Time for Two-stage Winner Designs with Different Endpoints when PFS ω_1=0.8 and $\omega_2 = 0.7$ and OS $\omega_1 = 0.8$

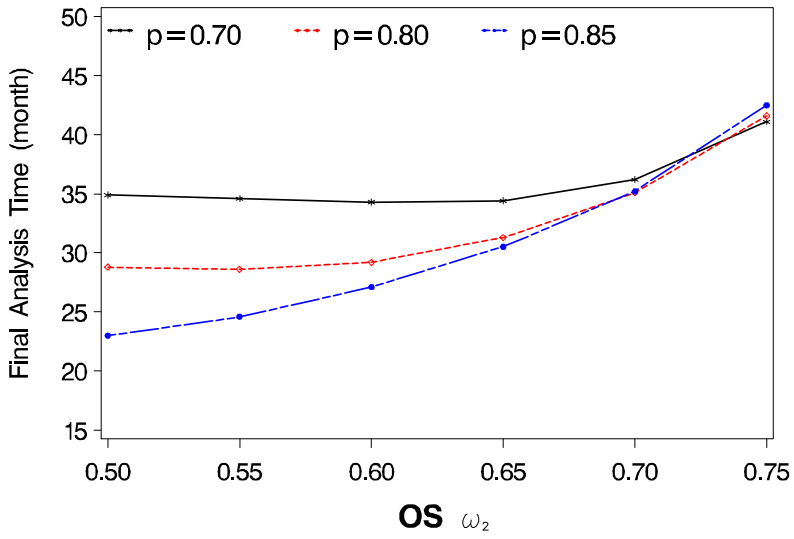

Figure 11.22: Final Analysis Time for Two-stage Winner Designs with Different Endpoints when PFS ω_1=0.8 and $\omega_2 = 0.7$ and OS $\omega_1 = 0.8$

The design parameters were also calculated when all the hazard ratios for OS and PFS were fixed under different winning probabilities (Table 11.8). Similar to the results when with the same endpoint was used at the interim and final analysis, the higher the winning probability was, the larger the number of PFS events at interim analysis is. Thus, more alpha spending penalty was needed to pay due to the interim analysis. The critical value for the final analysis becomes larger when the winning probability is increased.

The effects of PFS medians was also explored (see Table 11.9). When the hazard ratios were fixed, the total number of PFS events at the interim analysis was fixed. Only the time for interim analysis was affected by the median of PFS. The larger the PFS median was, the more time was required to reach the target number of PFS events.

Table 11.8: Two-stage Winner Designs with Different Winning Probabilities when $\omega_1 = 0.8$ and $\omega_2 = 0.7$ for Both OS and PFS

	Winning Probability				
Parameter	p=0.6	p=0.7	p=0.8	p=0.85	p=0.9
τ	0.0088	0.0517	0.1925	0.3537	0.6531
d_0	14.4	61.7	158.9	241.0	368.4
d_1	5.1	26.2	84.1	144.1	250.2
d_2	576.6	506.0	436.8	407.4	383.1
critical value	1.9770	1.9954	2.0176	2.0315	2.0484
interim time (months)	3.1	7.2	14.0	19.4	27.8
final time (months)	38.1	36.2	35.1	35.2	36.0

Table 11.9: Two-stage Winner Designs with Different PFS Medians When $\omega_1 = 0.8$ and $\omega_2 = 0.7$ for Both OS and PFS

PFS Median	Parameter	Winner Probability		
		p=0.7	p=0.8	p=0.85
4 month	τ	0.0865	0.2789	0.4716
	d_0	61.7	158.9	241.0
	d_1	43.2	119.8	189.5
	d_2	500.1	429.5	401.7
	critical value	2.0331	2.0764	2.1007
	interim time (months)	9.5	17.3	23.1
	final time (months)	36.7	35.7	35.9
6 month	τ	0.1175	0.3586	0.5848
	d_0	61.7	158.9	241.0
	d_1	57.9	150.0	228.9
	d_2	492.5	418.9	391.4
	critical value	2.0668	2.1253	2.1554
	interim time (months)	11.3	19.9	26.1
	final time (months)	36.9	36.0	36.1

11.2.2 Fix the information time approach

Another approach is to fix the information time τ. Instead of specifying a wining probability, an information time needs to be selected. When τ is decided, the critical value at the final analysis will be fixed given the type I error and the power requirement. One way to obtain the design parameters is to try different winning probabilities until one with desired information time is found. Another way is to calculate the parameters directly. For a given information time, the number of events in the winner and control groups can be calculated. Then the interim time, the event number in the two treatment arms at the interim and final analysis, can be determined iteratively.

Example (continued): Instead of using a fixed winning probability, the interim information time was fixed. Furthermore, it was assumed that all other parameters are the

same as in example 1. The information times selected are 0.3, 0.5, 0.7, and 1.0.

Results: The calculated parameters are presented in Tables11.10 and 11.11 and shown in Figures 11.23 to 11.28. Similar trends as shown in example 1 are observed here. When information time increases, the wining probability increases. The smaller the difference between two hazard ratios, the more the number of events at interim and final analysis are required. When $\tau = 1$, the two-stage winner design will have the largest winning probability at interim analysis.

Table 11.10: Two-stage Winner Designs Using Same Endpoints When $\omega_1 = 0.8$

		Information Time			
ω_2	Parameter	τ=0.3	τ=0.5	τ=0.7	τ=1
0.5	p	0.9073	0.9477	0.9708	0.9882
	d_0	31.8	47.7	64.9	92.8
	d_1	35.9	53.2	71.6	101.2
	d_2	119.7	106.4	102.3	101.2
	critical value	2.1334	2.1676	2.1901	2.2121
	interim time (months)	9.2	11.5	13.6	16.7
	final time (months)	17	16.5	16.6	16.7
0.55	p	0.8917	0.9348	0.9608	0.9821
	d_0	43.5	65.2	88.2	125.4
	d_1	48.8	72.0	96.5	135.6
	d_2	162.5	144.0	146.1	135.6
	critical value	2.1334	2.1676	2.1901	2.2121
	interim time (months)	10.7	13.4	15.9	19.6
	final time (months)	20.1	19.5	19.5	19.6
0.60	p	0.9	0.9145	0.9440	0.9705
	d_0	60.4	90.6	122.0	172.4
	d_1	67.0	99.2	132.2	184.7
	d_2	223.4	198.3	188.9	184.7
	critical value	2.1334	2.1676	2.1901	2.2121
	interim time (months)	12.6	15.9	19.0	23.4
	final time (months)	24.1	23.3	23.3	23.4
0.65	p	0.8304	0.8804	0.9137	0.9467
	d_0	84.7	128.5	172.5	241.7
	d_1	93.1	139.3	185.2	256.5
	d_2	310.3	278.5	264.5	256.5
	critical value	2.1334	2.1676	2.1901	2.2121
	interim time (months)	15.1	19.3	23.1	28.6
	final time (months)	29.3	28.6	28.6	28.6
0.70	p	0.7671	0.8183	0.8541	0.8928
	d_0	119.3	185.3	249.3	345.8
	d_1	129.7	198.7	264.7	363.3
	d_2	432.3	397.4	378.2	363.3
	critical value	2.1333	2.1676	2.1901	2.2121
	interim time (months)	18.2	23.8	28.8	35.9
	final time (months)	36.1	35.9	36.0	35.9
0.75	p	0.6616	0.7012	0.7307	0.7645
	d_0	166.8	267.6	363.0	498.9
	d_1	179	283.5	381.0	518.7
	d_2	596.8	566.9	544.3	518.7
	critical value	2.1334	2.1676	2.1901	2.2121
	interim time (months)	22.0	29.8	36.7	46.1
	final time (months)	44.8	45.7	46.3	46.1

Table 11.11: Two-stage Winner Designs with Same Endpoint under Different Information Time when $\omega_1 = 0.8$ and $\omega_2 = 0.7$

	Information Time								
	τ=0.2	τ=0.3	τ=0.4	τ=0.5	τ=0.6	τ=0.7	τ=0.8	τ=0.9	τ=1
p	0.7299	0.7671	0.7954	0.8183	0.8375	0.8541	0.8686	0.8814	0.8928
d_0	84.2	119.3	152.8	185.3	217.4	249.3	281.2	313.3	345.8
d_1	92.3	129.7	164.8	198.7	231.9	264.7	297.5	330.3	363.3
d_2	461.7	432.3	412.1	397.4	386.5	378.2	371.8	367.0	363.3
critical value	2.1082	2.1333	2.1524	2.1676	2.1799	2.1901	2.1987	2.206	2.2121
interim time (months)	14.8	18.2	21.1	23.8	26.4	28.8	31.2	33.5	35.9
final time (months)	36.5	36.1	35.9	35.9	35.9	36.0	36.0	36.0	35.9

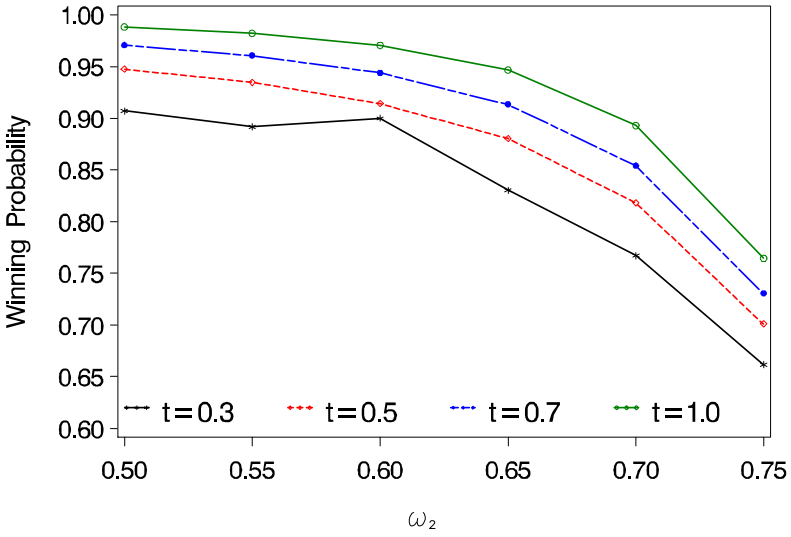

Figure 11.23: Winning Probabilities for Two-stage Winner Designs using Same Endpoint when $\omega_1 = 0.8$

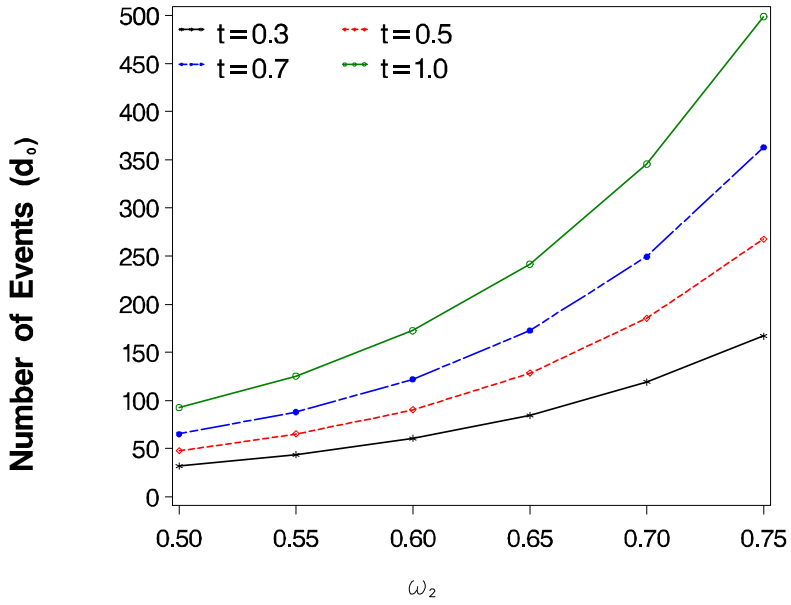

Figure 11.24: Total Events in Two Treatment Groups for Two-stage Winner Designs using Same Endpoint when $\omega_1 = 0.8$

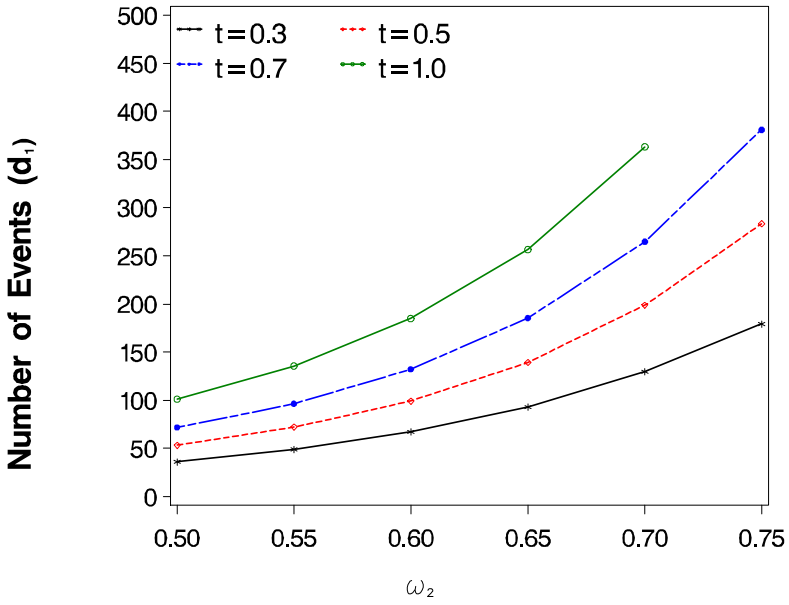

Figure 11.25: Total Events in Winner and Control Groups at First Stage for Two-stage Winner Designs using Same Endpoint when $\omega_1 = 0.8$

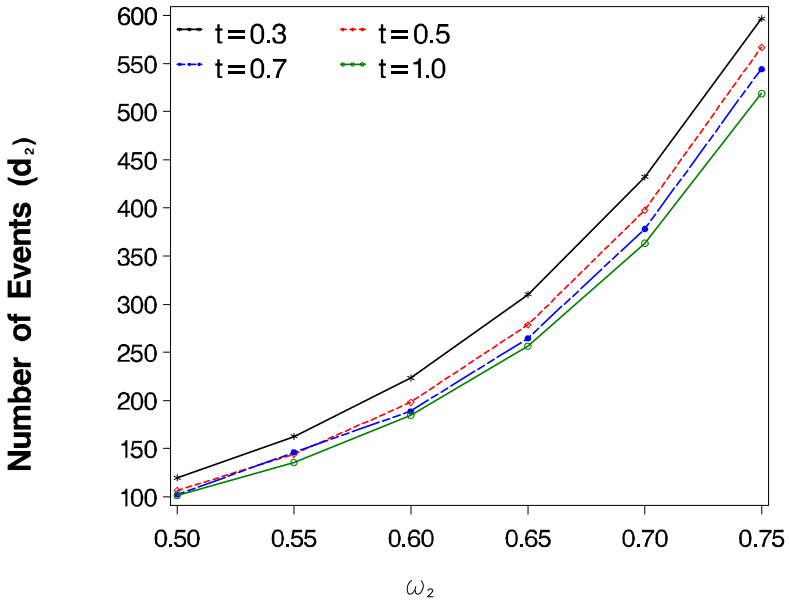

Figure 11.26: Total Events in Winner and Control Groups at Final Stage for Two-stage Winner Designs using Same Endpoint when $\omega_1 = 0.8$

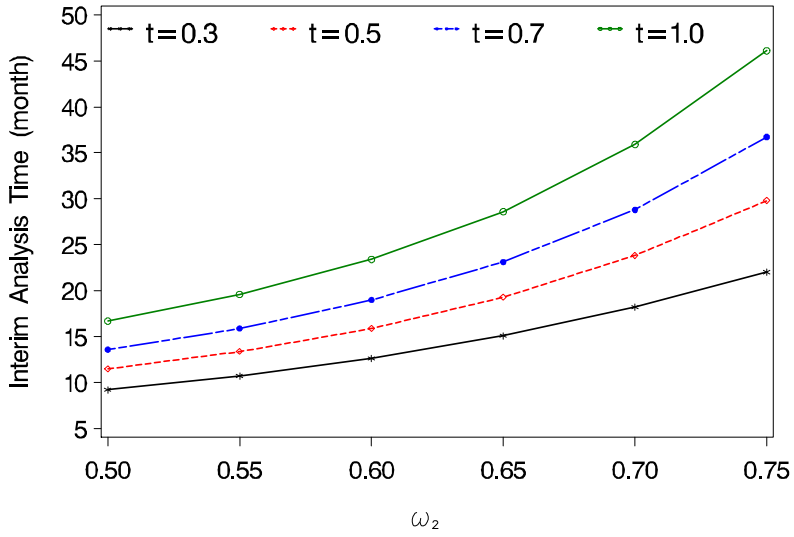

Figure 11.27: Interim Time for Two-stage Winner Designs using Same Endpoint when $\omega_1 = 0.8$

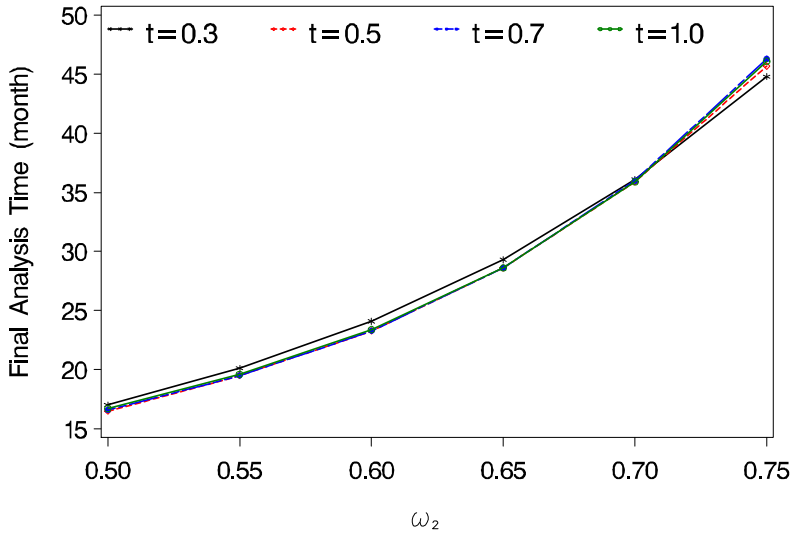

Figure 11.28: Final Time for Two-stage Winner Designs using Same Endpoint when $\omega_1 = 0.8$

11.2.3 Simulations

A. Same endpoint

Simulations were done to verify the calculations of winning probability, power, interim time, and the final time. To simulate the study, it was assumed that the accrual rate was 23 patients per month, and median survival for control was 7.5 months, $\omega_1 = 0.8$, and $\omega_2 = 0.7$. Various d_0 and d_2 were assumed under different design scenarios. All simulations were based on 10,000 repetitions.

As shown in Tables 11.12 and 11.13, the difference between the calculated results and the simulations results was very small. It is difficult to see the difference from Figures 11.29 to 11.34.

Table 11.12: Simulation Results for Two-stage Winner Designs with Same Endpoint under Fixed Winning Probabilty Approach

	Winning Probability							
	p=0.7		p=0.8		p=0.85		p=0.89279	
Parameter	calculation	simulation	calculation	simulation	calculation	simulation	calculation	simulation
winning probability	0.7000	0.6908	0.8000	0.8012	0.8500	0.8432	0.8928	0.8909
information time	0.1401	0.1373	0.4186	0.4149	0.6740	0.6685	1.0000	0.9962
power	0.9000	0.8978	0.9000	0.9012	0.9000	0.9013	0.9000	0.9003
d_1	68.2	66.9	171.2	169.7	256.2	254.7	363.3	362.6
interim time	12.4	12.4	21.6	21.6	28.2	28.1	35.9	35.8
final time	37.0	37.0	35.9	35.9	35.9	36.0	35.9	36.0

Table 11.13: Simulation Results for Two-stage Winner Designs with Same Endpoint under Fixed Information Time Approach

	Information Time							
	τ=0.3		τ=0.5		τ=0.7		τ=1	
Parameter	calculation	simulation	calculation	simulation	calculation	simulation	calculation	simulation
winning probability	0.7671	0.7729	0.8183	0.8221	0.8541	0.8524	0.8928	0.8909
information time	0.3000	0.2975	0.5000	0.4972	0.7000	0.6971	1.0000	0.9962
power	0.9000	0.9061	0.9000	0.9002	0.9000	0.9050	0.9000	0.9003
d_1	129.7	128.9	198.7	197.9	264.7	264.2	363.3	362.6
interim time	18.2	18.2	23.8	23.8	28.8	28.8	35.9	35.8
final time	36.1	36.2	35.9	36.0	36.0	36.1	35.9	36.0

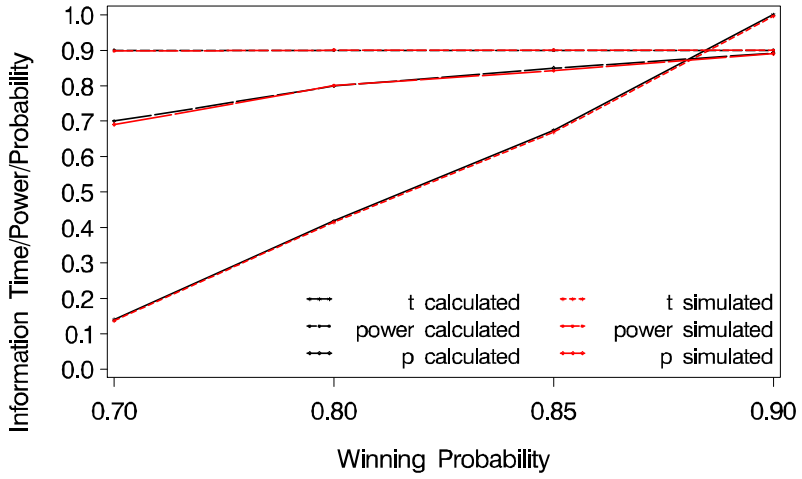

Figure 11.29: Comparison of Simulation Results for Information Time, Winning Probability, and Power in Fixed Winning Probability Approach

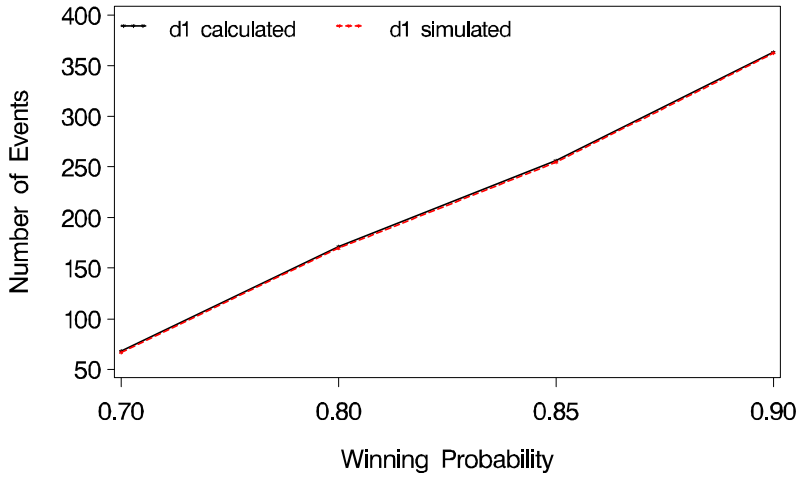

Figure 11.30: Comparison of Simulation Results for Total Events in Two Treatment Groups Power in Fixed Winning Probability Approach

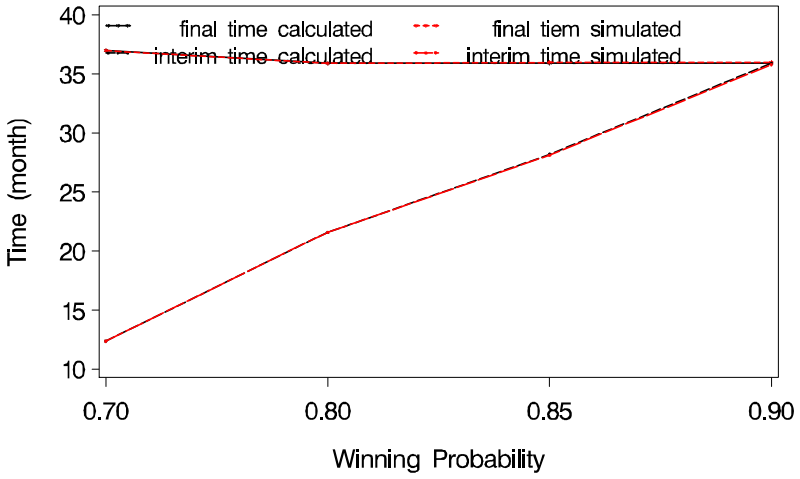

Figure 11.31: Comparison of Interim and Final Time in Fixed Winning Probability Approach

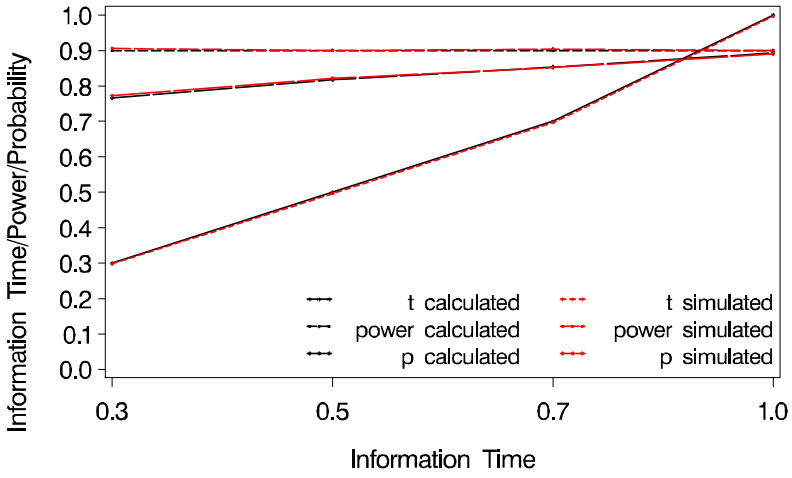

Figure 11.32: Comparison of Simulation Results for Information Time, Winning Probability, and Power in Fixed Information Time Approach

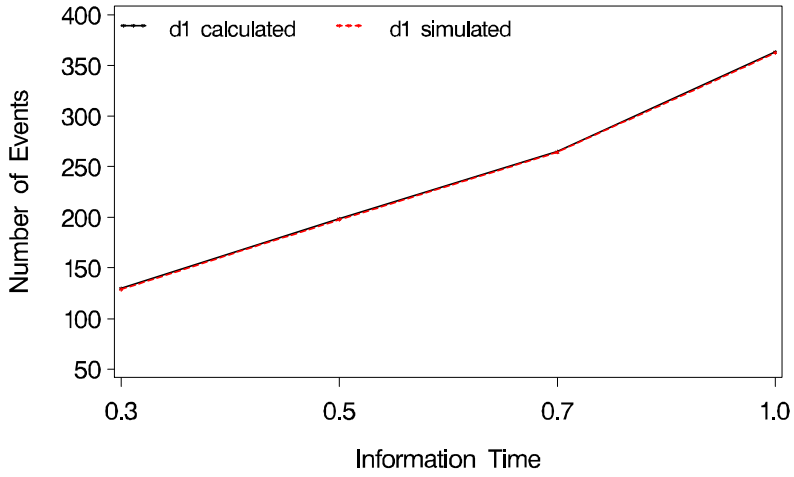

Figure 11.33: Comparison of Simulation Results for Total Events in Two Treatment Groups in Fixed Information Time Approach

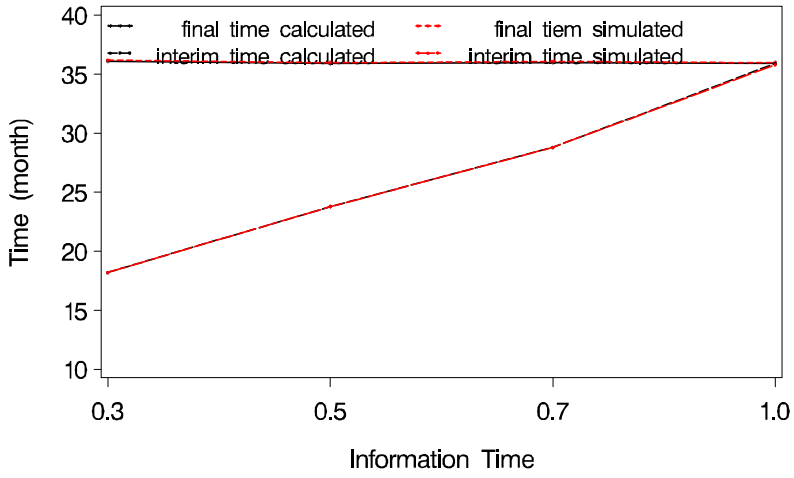

Figure 11.34: Comparison of Simulation Results for Interim and Final Time in Fixed Information Time Approach

B. Different endpoints

I) Bootstrap simulation plan

Step 1: Generate the original sample.

1. Assume uniform enrollment, randomly assign treatment arms (control, low, and high dose) with a 1:1:1 ratio
2. **Under the null hypothesis H_0:**
   (a) based on OS median, get the hazard rate (λ_2) for OS
   (b) based on PFS median, get the hazard rate (λ_1) for PFS
   (c) based on a) & b), get the hazard rate $(\lambda_0 = (\lambda_1 - \lambda_2))$ for T which is the recorded PFS
   (d) generate random sample for T for all treatment arms follow the same exponential distribution with parameter λ_0
   (e) generate random sample for OS for all treatment arms follow the same exponential distribution with parameter λ_2
   (f) generate random sample for PFS based on paired T and OS: $PFS = \min(T, OS)$
3. **Under the alternative hypothesis H_a:** repeat step 2 based on different medians for each arm
4. Interim time t_1 is determined by the total number d_0 of PFS events in two treatment groups at the interim.
5. Calculate PFS log-rank statistic to select a winner

6. Continue to assign the rest of the patients randomly to the winner and control arms
7. Final stopping time t_2 is determined by the total number of OS events
8. Calculate OS log-rank statistic LR at final analysis based on the original sample

Step 2: Bootstrap resampling (sample the same number of patients with replacement within each stage)

1. Bootstrap steps in stage 1
   (a) Based on those n_1 patients who enrolled in the 1st stage in the original sample, resample n_1 observations with replacement
   (b) randomly assign these observations to 3 treatment arms with 1:1:1 ratio
   (c) calculate PFS hazard ratio HR (low dose versus high dose)
   (d) pick a winner. If $HR > 1$ then pick high dose. If $HR < 1$ then pick low dose.
2. Bootstrap steps in stage 2
   (a) Based on the rest of the n_2 patients who enrolled in the 2nd stage in the original sample, resample n_2 observations with replacement
   (b) Randomly assign these observations to the winner and control arms with 1:1 ratio
   (c) Combine observations that assigned to the winner and control arms in the stage 1 and stage 2
   (d) Calculate OS log-rank statistic for the bootstrap sample

Step 3: Repeat Step 2 for 1000 time

Step 4: Determine the a critical value or a p-value based on the log-rank statistics from step 2

1. Critical value will be the 95th percentile of all log-rank statistics
2. P-value is $p = \sum_{i=1}^{1000} I_{(LR_i \geq LR)}/1000$, where LR_i is OS log rank statistic from i^{th} bootstrap sample.

Step 5: Repeat Step 1 to 4 for 1000 times to determine the overall critical value and bootstrap power.

1. Overall critical value is $c = \sum_{i=1}^{1000} c_i/1000$, where c_i is the critical value determined in Step 4 in the i^{th} repetition.
2. Resampling power is $1 - \beta = \sum_{i=1}^{1000} I_{(LR \geq c_i)}/1000$

II) Simulation results

Summary of the simulation results for critical values are presented in Table 11.14. The average critical values from the simulations were compared with the calculated values. As shown in Table 11.15, the calculated values and simulated values are very close to each other.

Table 11.14: Simulation Results for Two-stage Winner Designs with Different Endpoint when PFS ω_1=0.8 and $\omega_2 = 0.7$ and OS $\omega_1 = 0.8$

OS ω_2	Critical Value	Winner Probability		
		p=0.7	p=0.8	p=0.85
0.50	min	1.7275	1.7985	1.7711
	p25	1.9414	1.9744	2.0100
	p50	1.9991	2.0347	2.0658
	p75	2.0615	2.0893	2.1283
	max	2.2596	2.2885	2.3412
	mean	2.0018	2.0353	2.0686
	std	0.0874	0.0856	0.0859
	*power**	0.8990	0.9090	0.9110
0.70	min	1.7550	1.7195	1.7881
	p25	1.9445	1.9676	1.9721
	p50	1.9948	2.0291	2.0285
	p75	2.0557	2.0789	2.0888
	max	2.2880	2.3120	2.3696
	mean	2.0006	2.0263	2.0320
	std	0.0838	0.0836	0.0880
	*power**	0.9140	0.9400	0.9440

* nominal=0.90

Table 11.15: Comparison of Simulation Results for Two-stage Winner Designs with Different Endpoint when PFS $\omega_1 = 0.8$ and $\omega_2 = 0.7$ and OS $\omega_1 = 0.8$

OS ω_1	Parameter	Winning Probability					
		p=0.7		p=0.8		p=0.85	
		calculated	simulated	calculated	simulated	calculated	simulated
0.50	critical value	1.9663	2.0018	2.0271	2.0353	2.0597	2.0686
0.70	critical value	1.9954	2.0006	2.0176	2.0263	2.0315	2.0320

Chapter 12

Conclusion

In Chapter 10, the correlation between the log-rank test statistics based on either the same or different survival endpoints used at the interim and final analysis were derived. The covariance matrix for log-rank test statistics were identified. Since the log-rank test statistics at the interim and final analysis were asymptotically normally distributed, the statistical framework for the two-stage winner design discussed by Lan et al [19, 35] could be extended to the trails with survival outcomes.

In Chapter 11, the two-stage winner design parameters were calculated in several scenarios. In practice, before choosing a two-stage winner design as the study design, one needs to do a best estimation of design parameters: how big is the treatment difference between the two treatment arms? Are there any time requirements to complete the whole study? Is there a patient or a resource limitation? A higher winning probability will ensure efficacious treatment can be selected to the final stage and reduce exposure to inferior treatment. When treatment difference between two treatment arms is fixed, more sample

size is required to obtain a higher winning probability. The larger the difference between two treatment arms is, the higher winning probability can be achieved by using a limited sample size. However the winning probability will reach its maximum value when the information time is approaching to 1. When a surrogate endpoint is used at the interim analysis, the time for interim analysis is determined by the treatment effect of surrogate endpoint. When the magnitude of surrogate treatment effect is larger than the treatment effect of primary endpoint, only a limited number of surrogate events are required at the interim. However, when the magnitude of surrogate treatment effect is smaller, the required event number for surrogate endpoint will increase dramatically. So an appropriate selected surrogate endpoint is crucial in two-stage winner design.

Chapter 13

Discussion and Future Works

13.1 Discussion

In part I of this dissertation, the expected sample size for the two-stage sample size re-estimation design was derived. Performance of adaptive designs were measured and compared. In part II of this dissertation, asymptotic correlation of log-rank statistics in two-stage winner design while using the same endpoint or using different endpoints at interim and final analysis was derived. Design features were studied and compared.

The most interesting conclusion in part I is that 5-looks unequal-spaced GS designs with HP01 and HP005 boundaries can achieve similar or better performance than sample size re-estimation designs. Especially when the total number of looks for re-estimation designs is small, performance improvement is prominent. While group sequential designs are well accepted by regulatory agencies, other types of adaptive designs like sample size re-estimation designs are still not fully understood and accepted. Thus, this finding becomes very crucial under the current FDA guidance on adaptive design use in clinical

trials. Pharmaceutical companies can use group sequential trials to achieve similar performance as sample size re-estimation trials by simply adding additional interim analysis, using selected type of boundaries, and adjusting the patient increment type between interim looks. Adding more interim looks in a clinical trial may require additional work on the trial operation side. However, it will become trivial, if we can save resources and less patients can be exposed to non-efficacious treatments.

Using interim analysis to do treatment selection becomes popular. However, some efficacy endpoints, such as death, may not be a good choice for quick efficacy analysis at the interim analysis. Thus, using surrogate endpoints should be necessary. Theoretical work for two-stage winner design using continuous endpoints was developed by Shun, Lan, and Soo. It is very challenging to derive the correlation of log-rank test statistics when different survival endpoints are used at interim and final analyses. The correlation is crucial in Type I error adjustment and boundary calculations. Thus, derivation of asymptotic correlation of log-rank statistics is the key item in part II. All other two-stage winner design features are studied based on the results of this asymptotic correlation.

13.2 Future Work

Expected sample size for weighted two-stage sample size re-estimation designs are developed in part I. However, because there are more assumptions required, the derivation for unweighted designs was not done. Thus, this needs to be further explored. Also the expected sample size was derived for re-estimation designs with total 2 looks. It could be generalized to designs with multiple looks.

In part II of this dissertation, a two-stage winner design was studied when considering a uniform patient enrollment and assumed no censoring process. As future work, instead of uniform distribution, other patient enrollment functions can be considered. A more complicated design with a censoring process may make two-stage winner design more practical. Thus, this could be explored in the future.

Appendix A

Conditional Power C_P for Weighted Method (CHW)

Notation:

N : Final sample size

n_L and n_k: Planned one arm sample size at L$^{\text{th}}$ and k$^{\text{th}}$ look, L<k

t_L and t_k : Information fraction for L$^{\text{th}}$ look ($\frac{n_L}{N}$) and k$^{\text{th}}$ look ($\frac{n_k}{N}$).

c_k: Boundary at k$^{\text{th}}$ look

Z_L and Z_k : Z-score at L$^{\text{th}}$ and k$^{\text{th}}$ look

δ : Assumed treatment effect

Since

$$\begin{aligned} C_P &= P(Z_k \geq c_k | Z_L) \\ &= P\left(\sqrt{\frac{n_L}{N}} Z_L + \sqrt{1 - \frac{n_L}{N}} Z_k \geq c_k | Z_L\right) \end{aligned}$$

$$= P\left(Z_k \geq \frac{c_k - \sqrt{\frac{n_L}{N}}Z_L}{\sqrt{1-\frac{n_L}{N}}}|Z_L\right)$$
$$= P\left(Z_k \geq \frac{c_k - \sqrt{\frac{n_L}{N}}Z_L}{\sqrt{1-\frac{n_L}{N}}}\right)$$
$$= P\left(Z_k \geq \frac{c_k - \sqrt{t_L}Z_L}{\sqrt{1-t_L}}\right)$$

where

$$Z_k = \frac{\sum_{i=1}^{n_k}(x_i - y_i)}{n_k} \sim N\left(\sqrt{\frac{n_k}{2}}\delta, 1\right),$$

we have

$$C_P = 1 - \Phi\left[\frac{c_k - \sqrt{t_L}Z_L}{\sqrt{1-t_L}} - \sqrt{\frac{N-n_L}{2}}\delta\right]$$
$$= 1 - \Phi\left[\frac{c_k - \sqrt{t_L}Z_L - \sqrt{1-t_L}\sqrt{\frac{N-n_L}{2}}\delta}{\sqrt{1-t_L}}\right].$$

For $L = 1$ and $k = 2$,

$$C_P = 1 - \Phi\left[\frac{c_2 - \sqrt{t_1}Z_1 - \sqrt{1-t_1}\sqrt{\frac{n_2}{2}}\delta}{\sqrt{1-t_1}}\right],$$

hence $C_P \geq C_{P_0}$,

$$1 - \Phi\left[\frac{c_2 - \sqrt{t_1}Z_1 - \sqrt{1-t_1}\sqrt{\frac{n_2}{2}}\delta}{\sqrt{1-t_1}}\right] \geq C_{P_0},$$

and

$$\frac{c_2 - \sqrt{t_1}Z_1 - \sqrt{1-t_1}\sqrt{\frac{n_2}{2}}\delta}{\sqrt{1-t_1}} \leq Z_{1-C_{P_0}},$$

and

$$\frac{c_2 - \sqrt{t_1}Z_1 - \sqrt{1-t_1}\sqrt{\frac{n_2}{2}}\delta}{\sqrt{1-t_1}} \leq -Z_{C_{P_0}},$$

therefore

$$Z_1 \geq \frac{\sqrt{1-t_1}Z_{C_{P_0}} - \sqrt{1-t_1}\sqrt{\frac{n_2}{2}}\delta + c_2}{\sqrt{t_1}}.$$

Appendix B

Sample Size and Power Curves for Adaptive Designs with Treatment Effect Follow Uniform Distribution

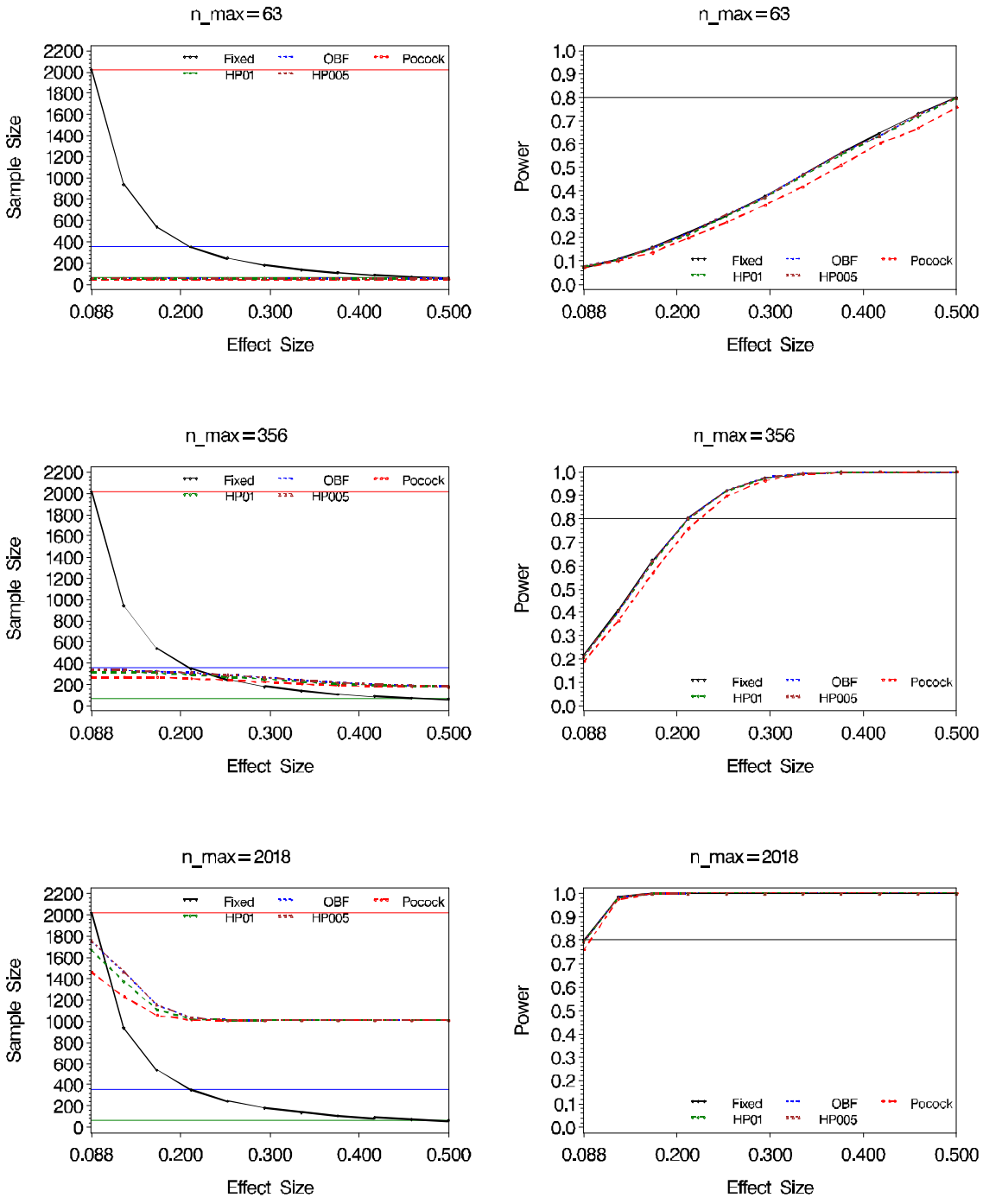

Figure B.1: Sample Size and Power Curves for Two-look Group Sequential Designs when Treatment Effect Follows Uniform Distribution

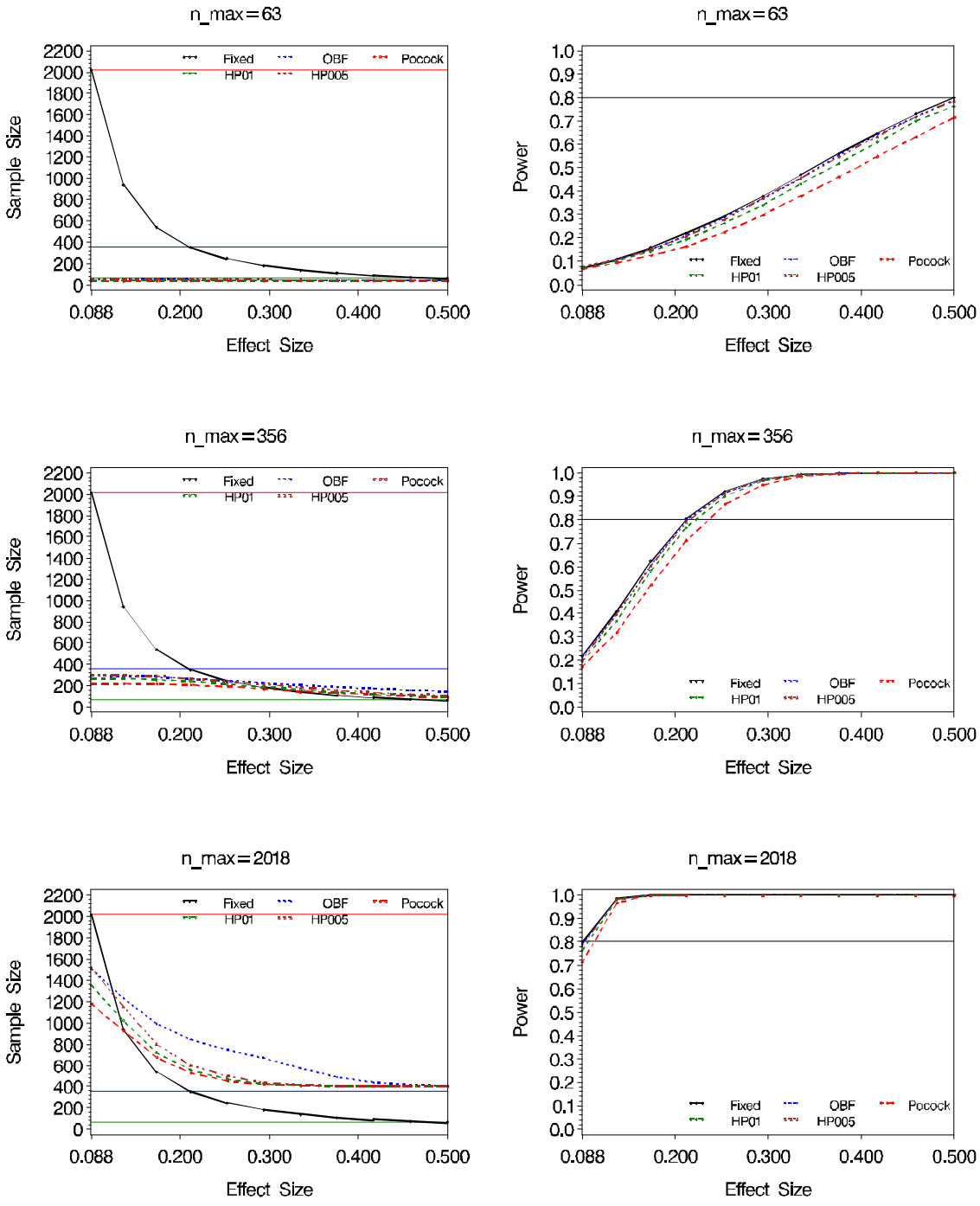

Figure B.2: Sample Size and Power Curves for Five-look Equal-spaced Group Sequential Designs when Treatment Effect Follows Uniform Distribution

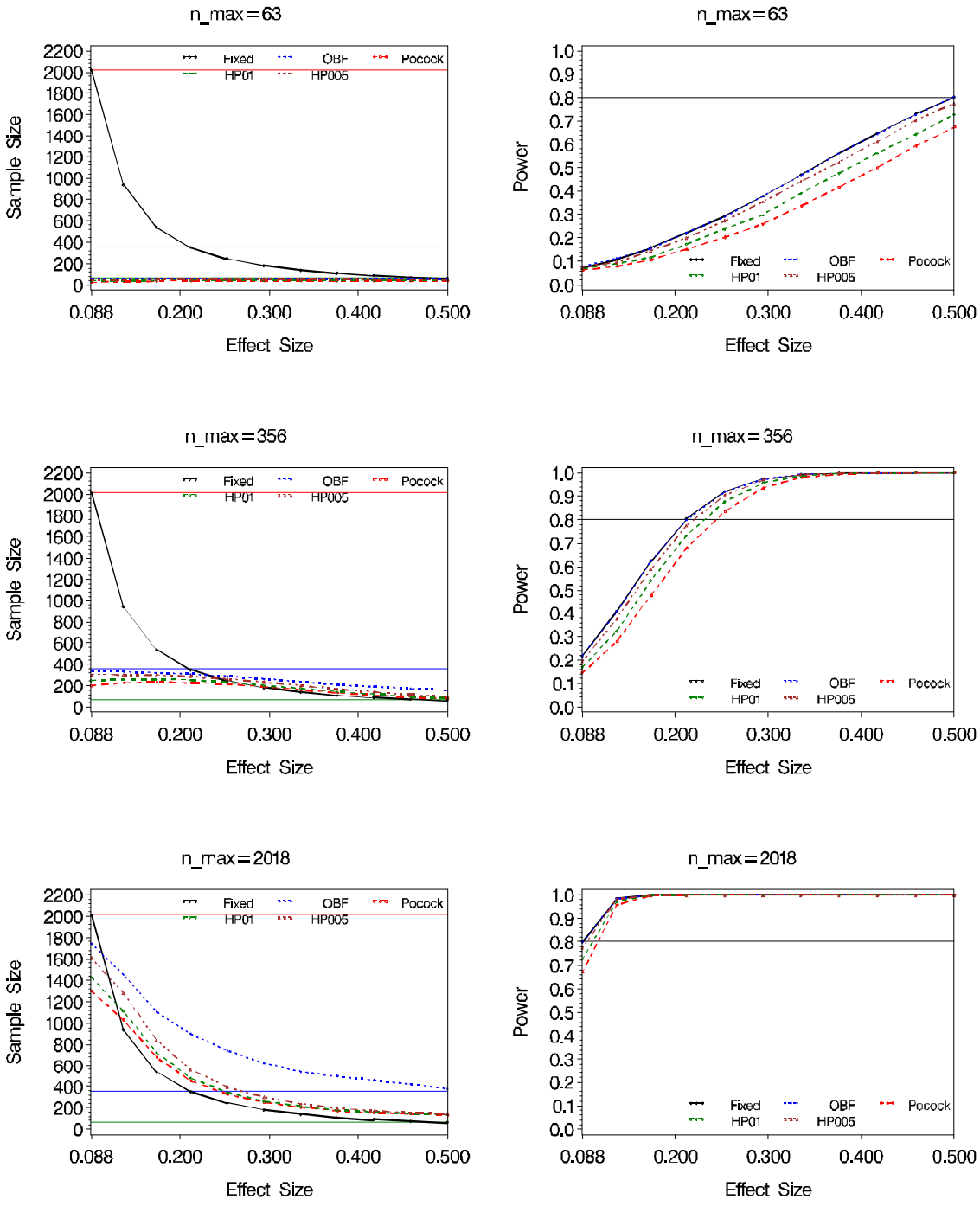

Figure B.3: Sample Size and Power Curves for Five-look Unequal-spaced Group Sequential Designs when Treatment Effect Follows Uniform Distribution

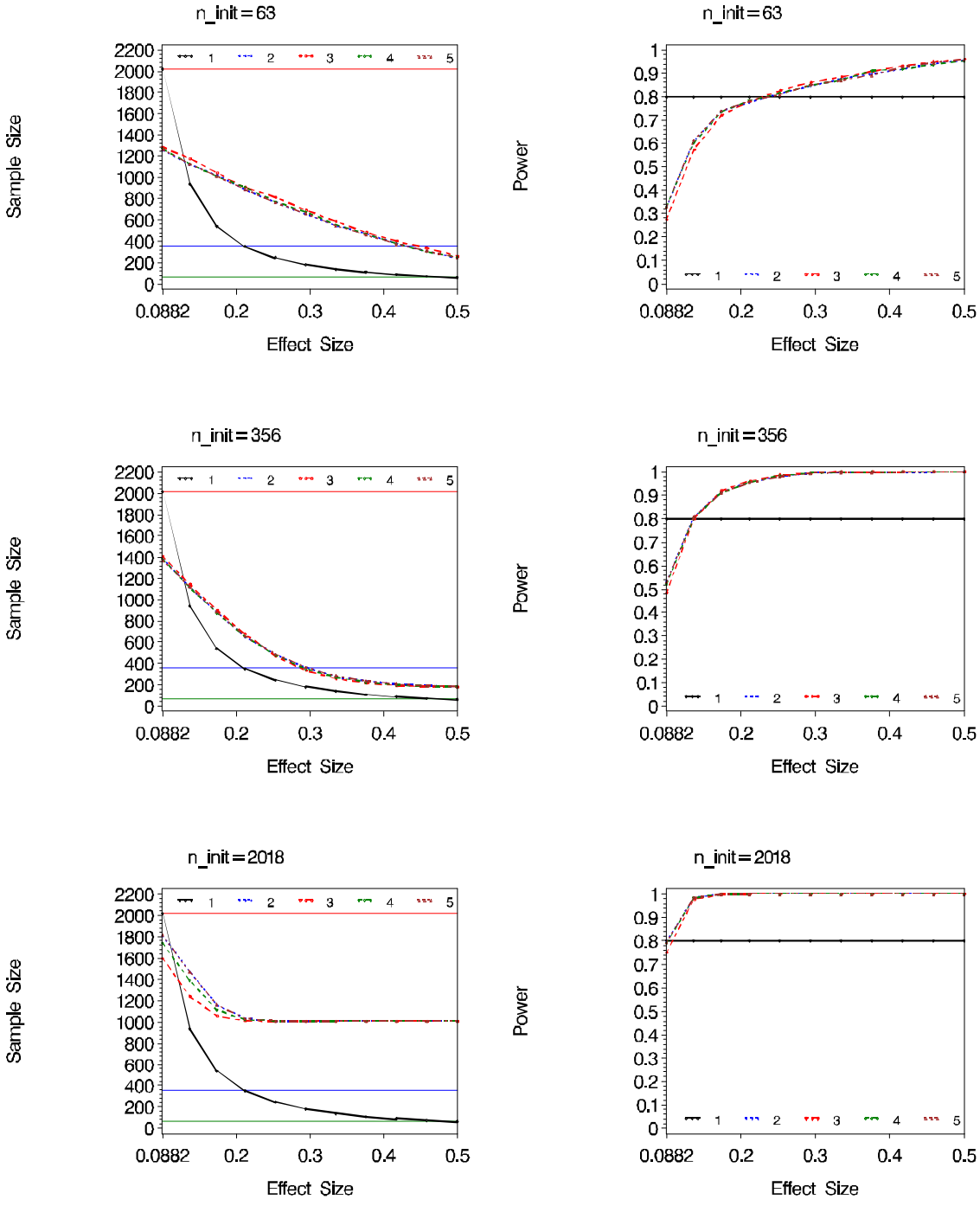

Figure B.4: Sample Size and Power Curves for Two-look Weighted Sample Size Re-estimation Designs when Treatment Effect Follows Uniform Distribution (re-estiamtion at look 1)

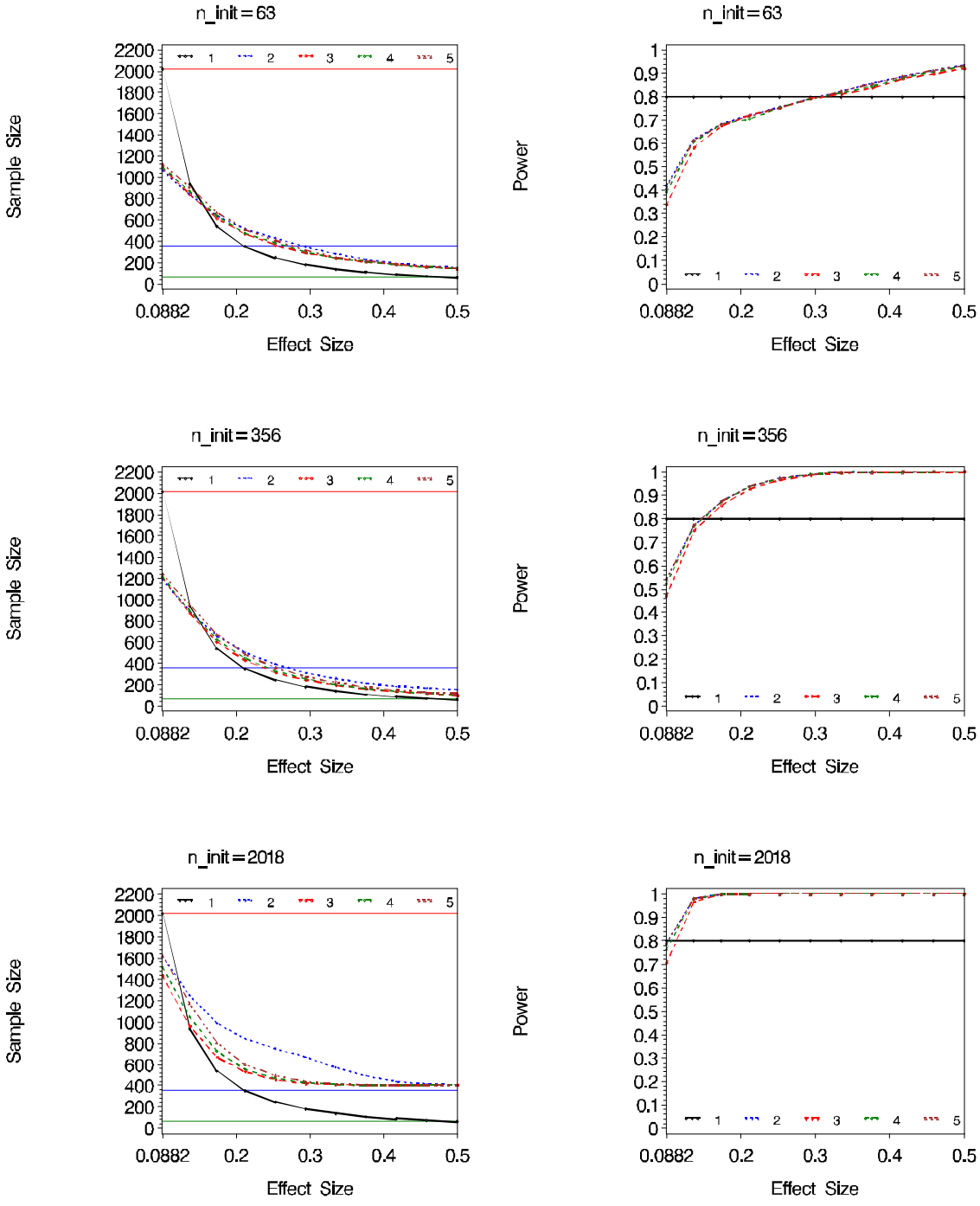

Figure B.5: Sample Size and Power Curves for Five-look Equal-spaced Weighted Sample Size Re-estimation Designs when Treatment Effect Follows Uniform Distribution (re-estiamtion at look 1)

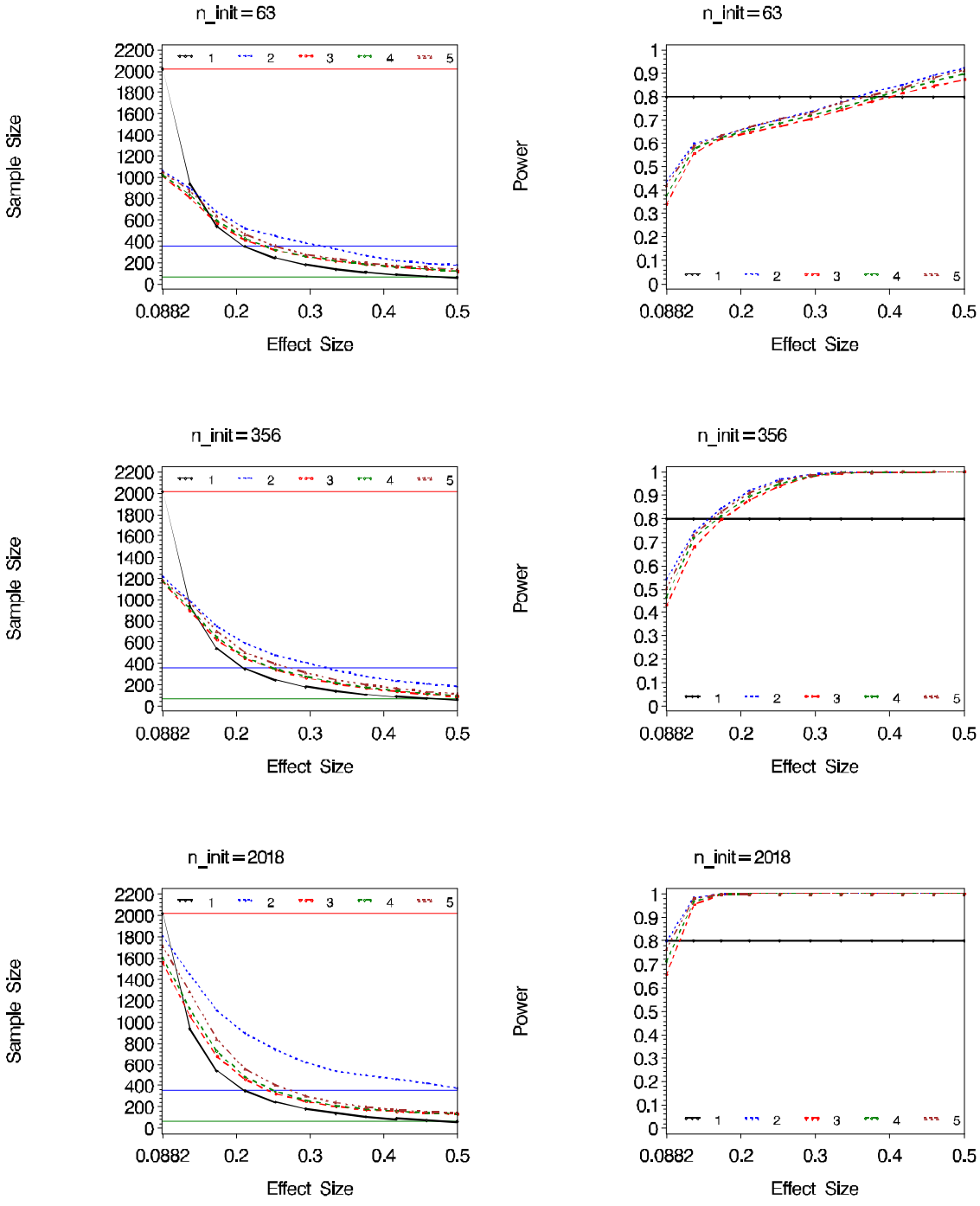

Figure B.6: Sample Size and Power Curves for Five-look Unequal-spaced Weighted Sample Size Re-estimation Designs when Treatment Effect Follows Uniform Distribution (re-estiamtion at look 1)

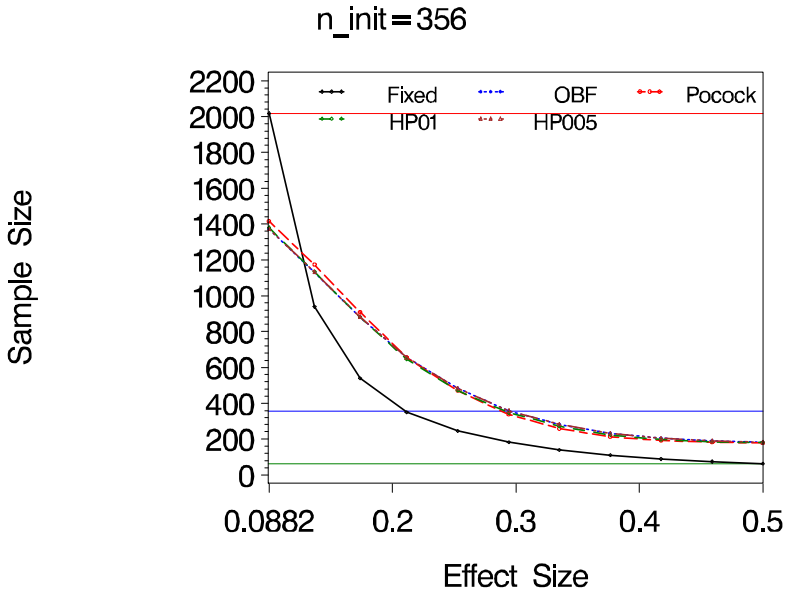

Figure B.7: Sample Size Curves for Two-look Unweighted Sample Size Re-estimation Designs when Treatment Effect Follows Uniform Distribution

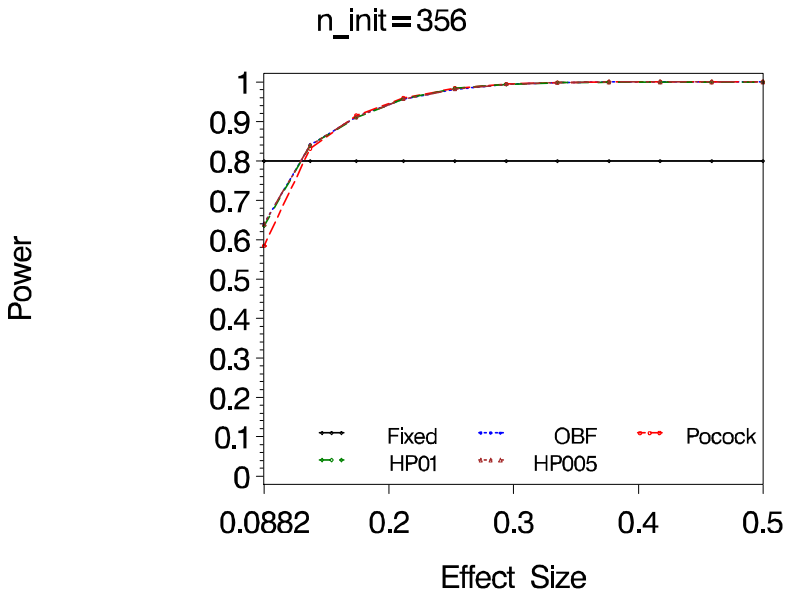

Figure B.8: Power Curves for Two-look Unweighted Sample Size Re-estimation Designs when Treatment Effect Follows Uniform Distribution

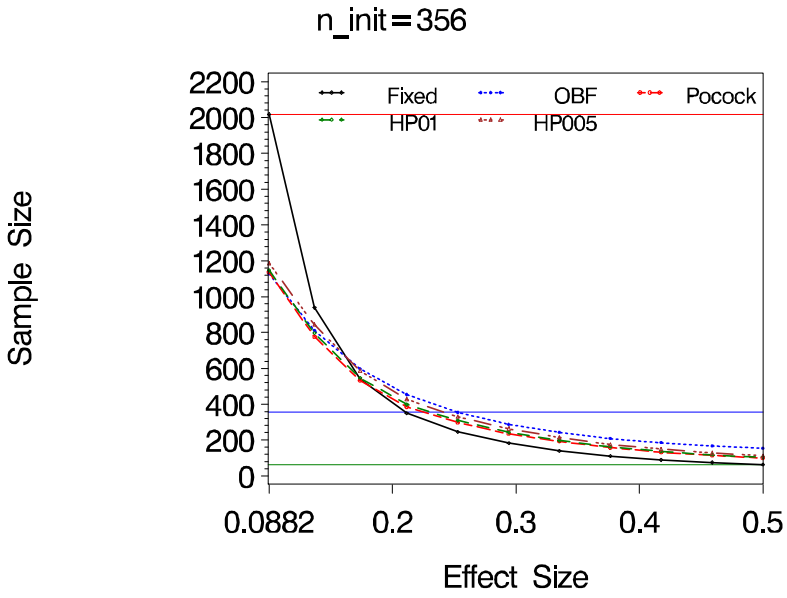

Figure B.9: Sample Size Curves for Five-look Equal-spaced Unweighted Sample Size Re-estimation Designs when Treatment Effect Follows Uniform Distribution (re-estiamtion at look 1)

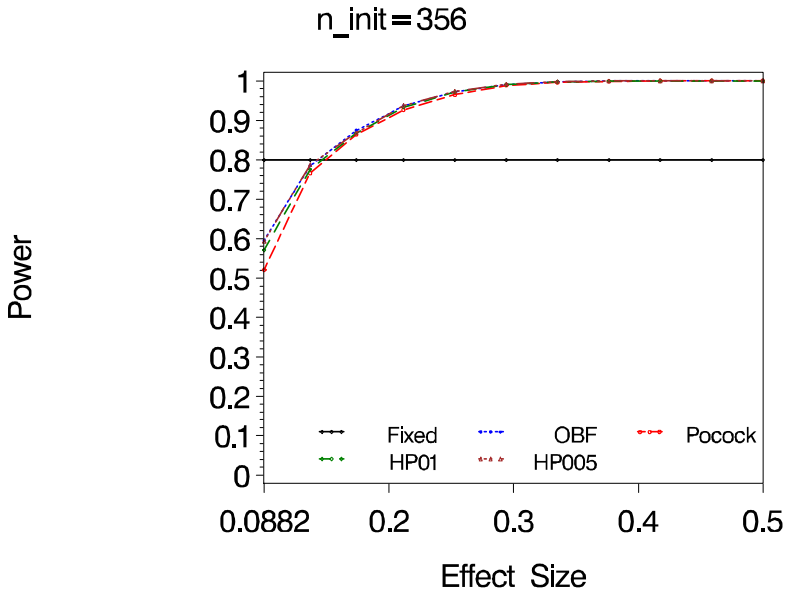

Figure B.10: Power Curves for Five-look Equal-spaced Unweighted Sample Size Re-estimation Designs when Treatment Effect Follows Uniform Distribution (re-estiamtion at look 1)

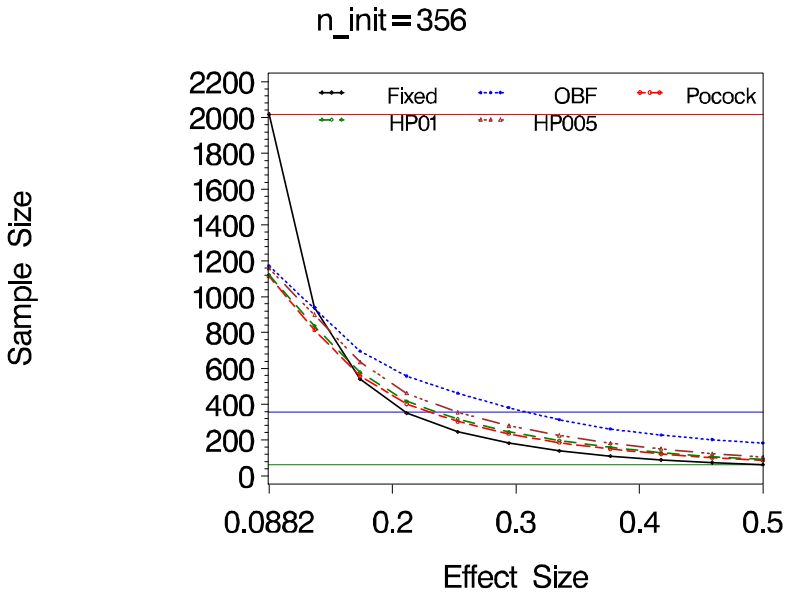

Figure B.11: Sample Size Curves for Five-look Unequal-spaced Unweighted Sample Size Re-estimation Designs when Treatment Effect Follows Uniform Distribution (re-estiamtion at look 1)

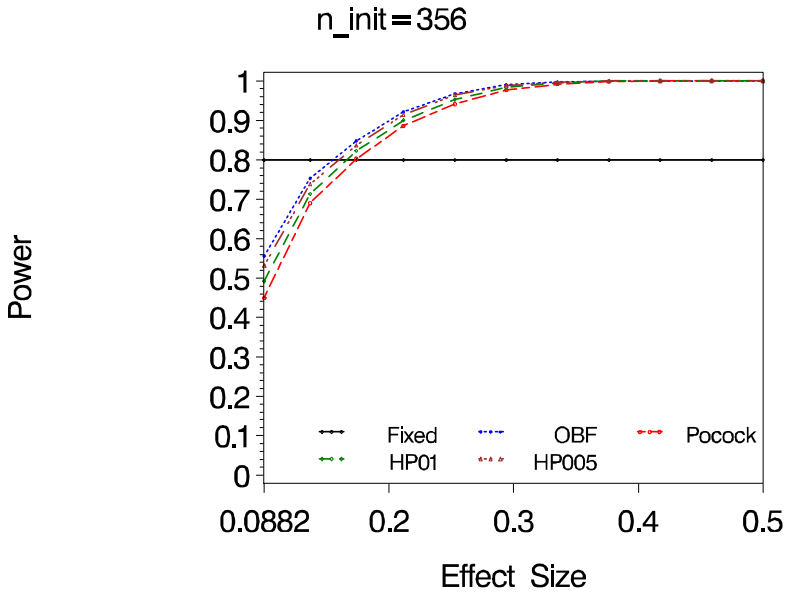

Figure B.12: Power Curves for Five-look Unequal-spaced Unweighted Sample Size Re-estimation Designs when Treatment Effect Follows Uniform Distribution (re-estiamtion at look 1)

Appendix C

Sample Size and Power Curves for Adaptive Designs with Treatment Effect Follow Beta Distribution

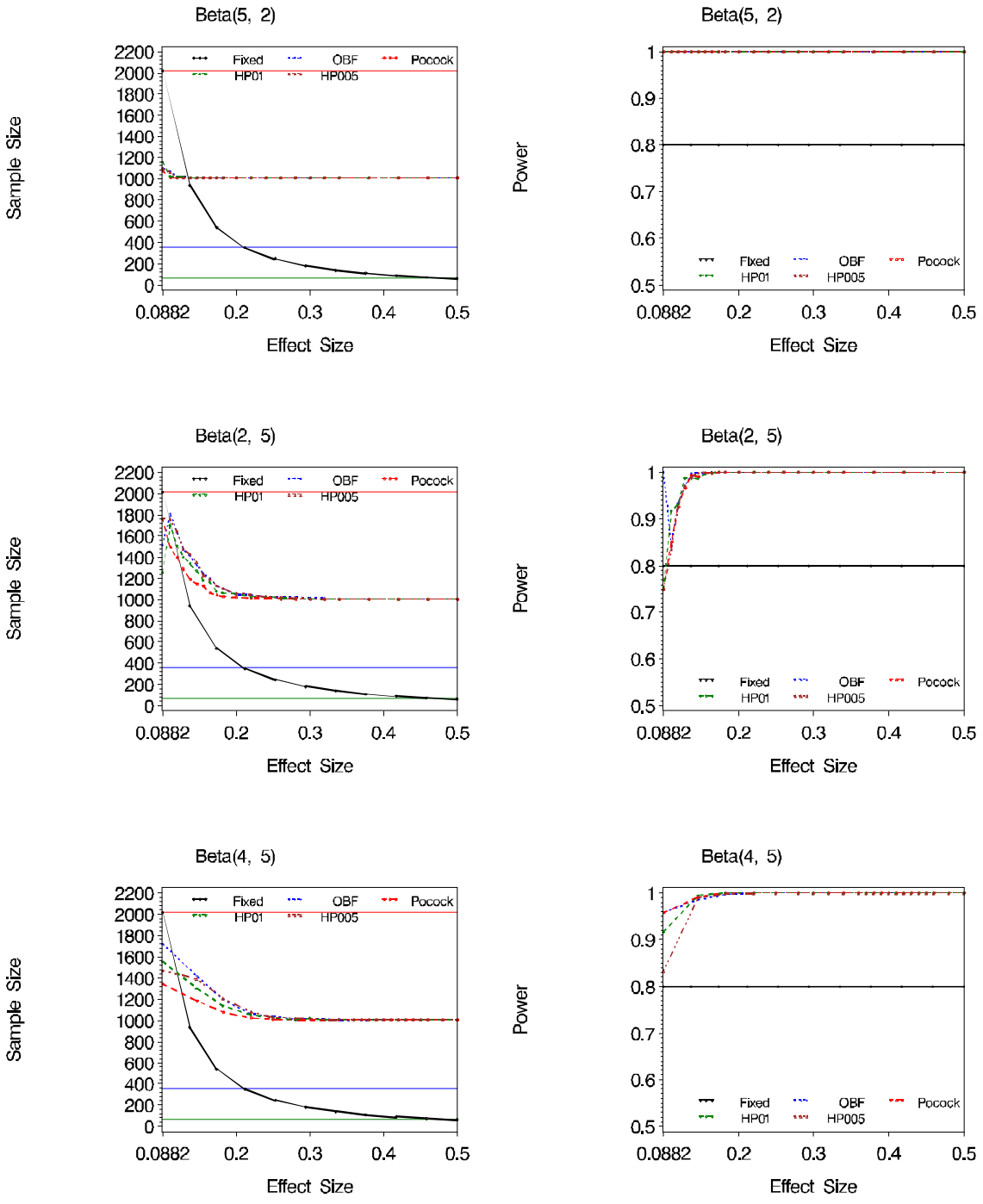

Figure C.1: Sample Size and Power Curves for Two-look Group Sequential Designs when Treatment Effects follow Beta Distributions

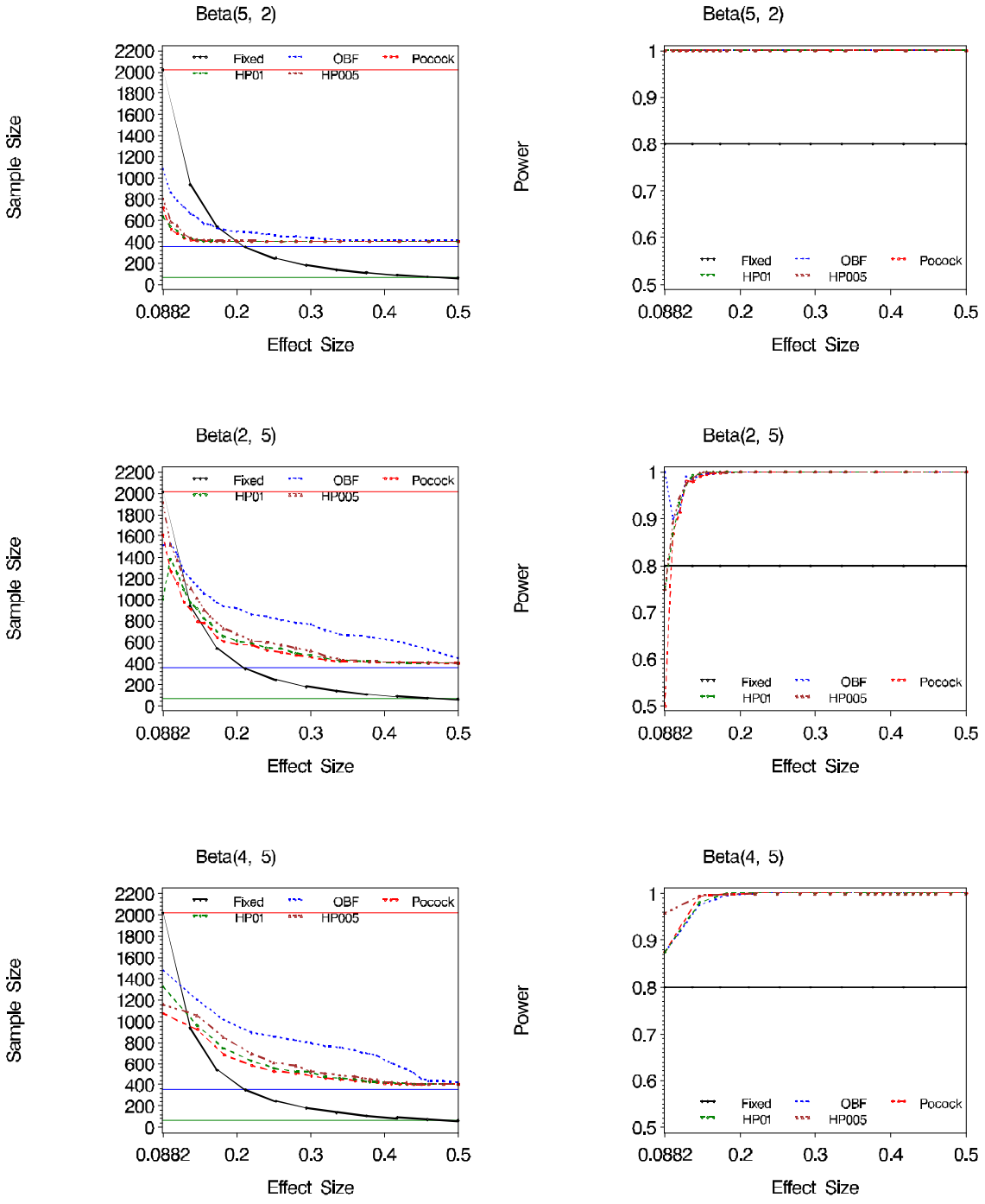

Figure C.2: Sample Size and Power Curves for Five-look Equal-spaced Group Sequential Designs when Treatment Effects follow Beta Distributions

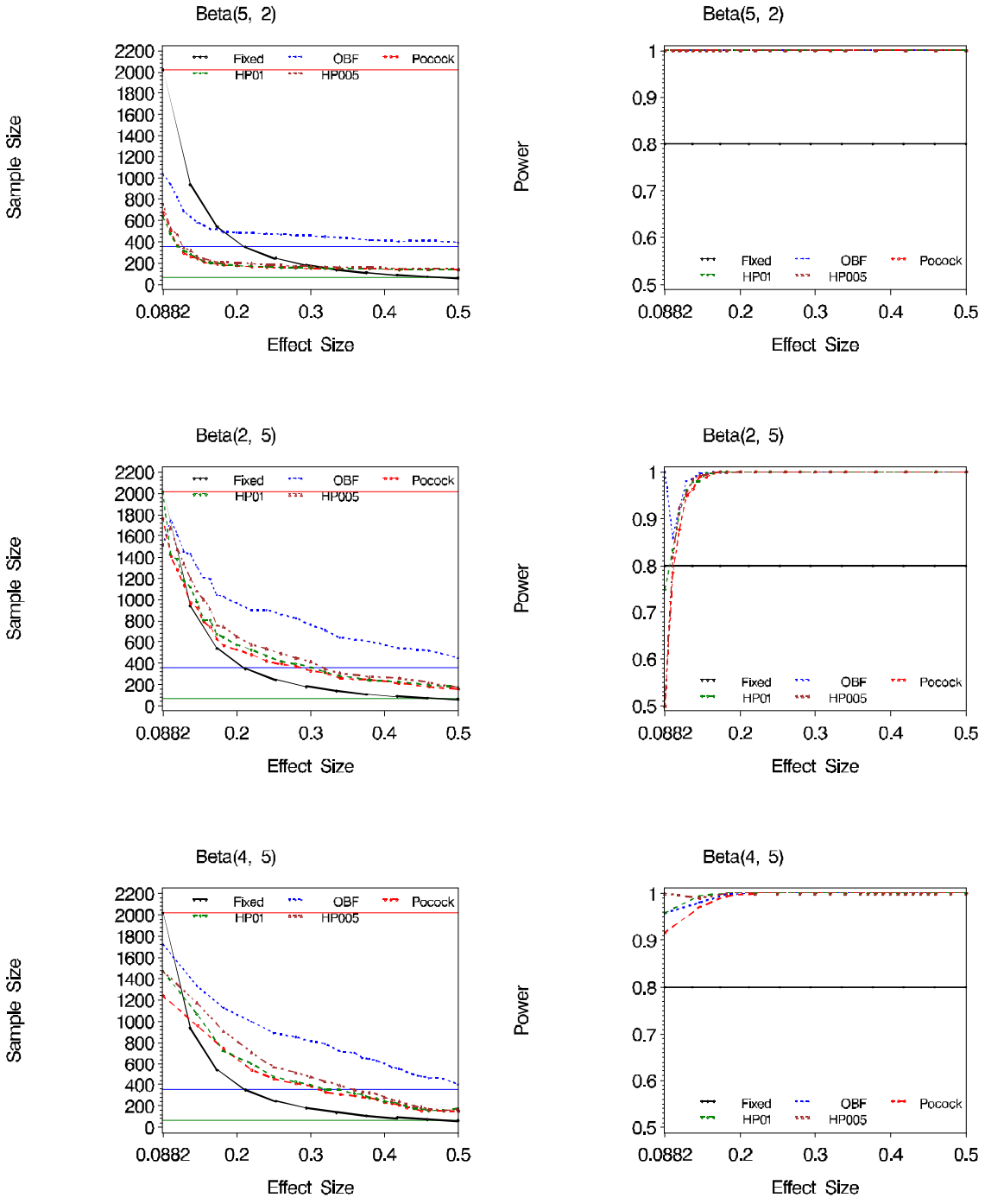

Figure C.3: Sample Size and Power Curves for Five-look Unequal-spaced Group Sequential Designs when Treatment Effects follow Beta Distributions

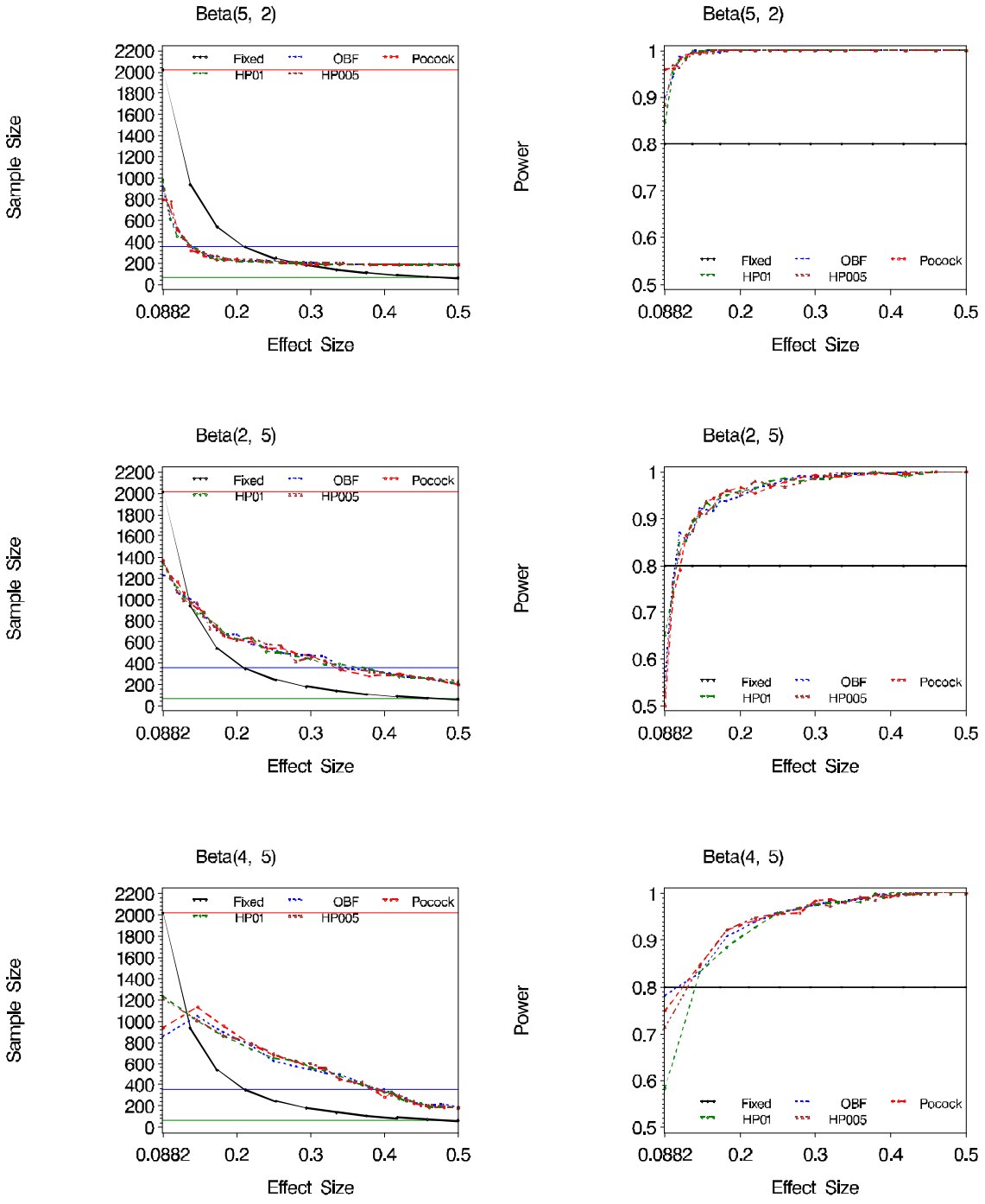

Figure C.4: Sample Size and Power Curves for Two-look Weighted Sample Size Re-estimation Designs when Treatment Effects follow Beta Distributions

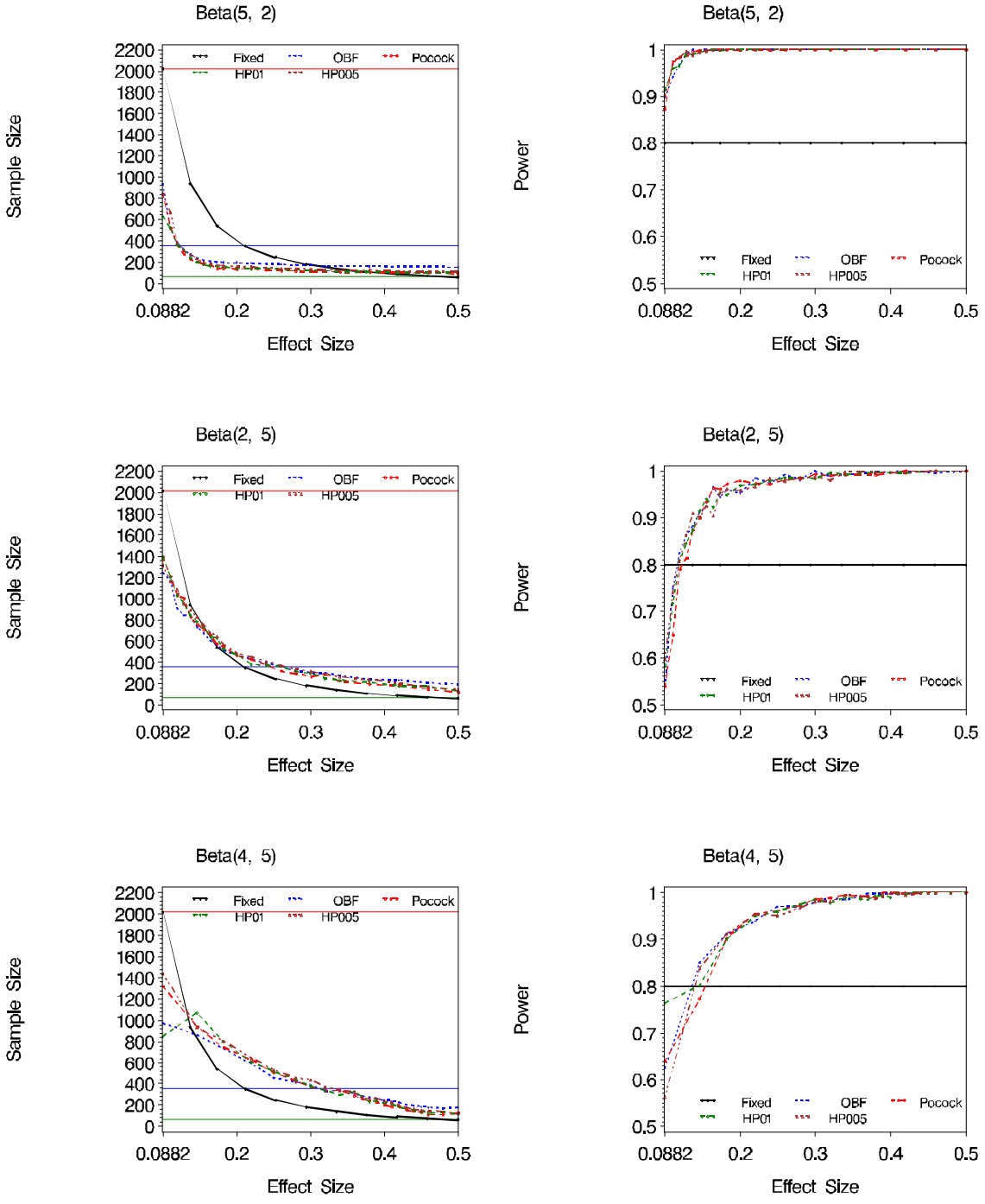

Figure C.5: Sample Size and Power Curves for Five-look Equal-spaced Weighted Sample Size Re-estimation Designs when Treatment Effects follow Beta Distributions (re-estimation at look 3)

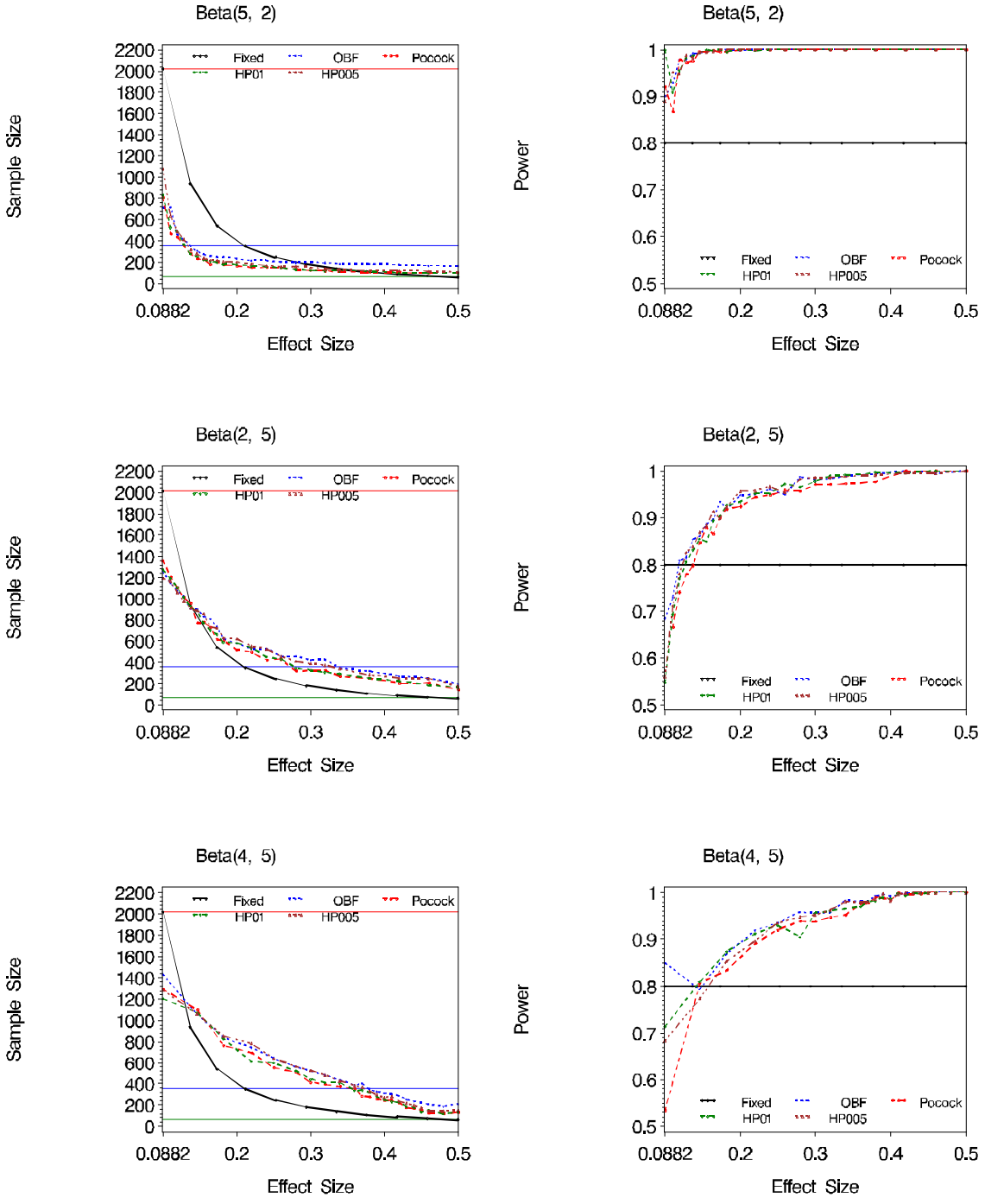

Figure C.6: Sample Size and Power Curves for Five-look Unequal-spaced Weighted Sample Size Re-estimation Designs when Treatment Effects follow Beta Distributions (re-estimation at look 3)

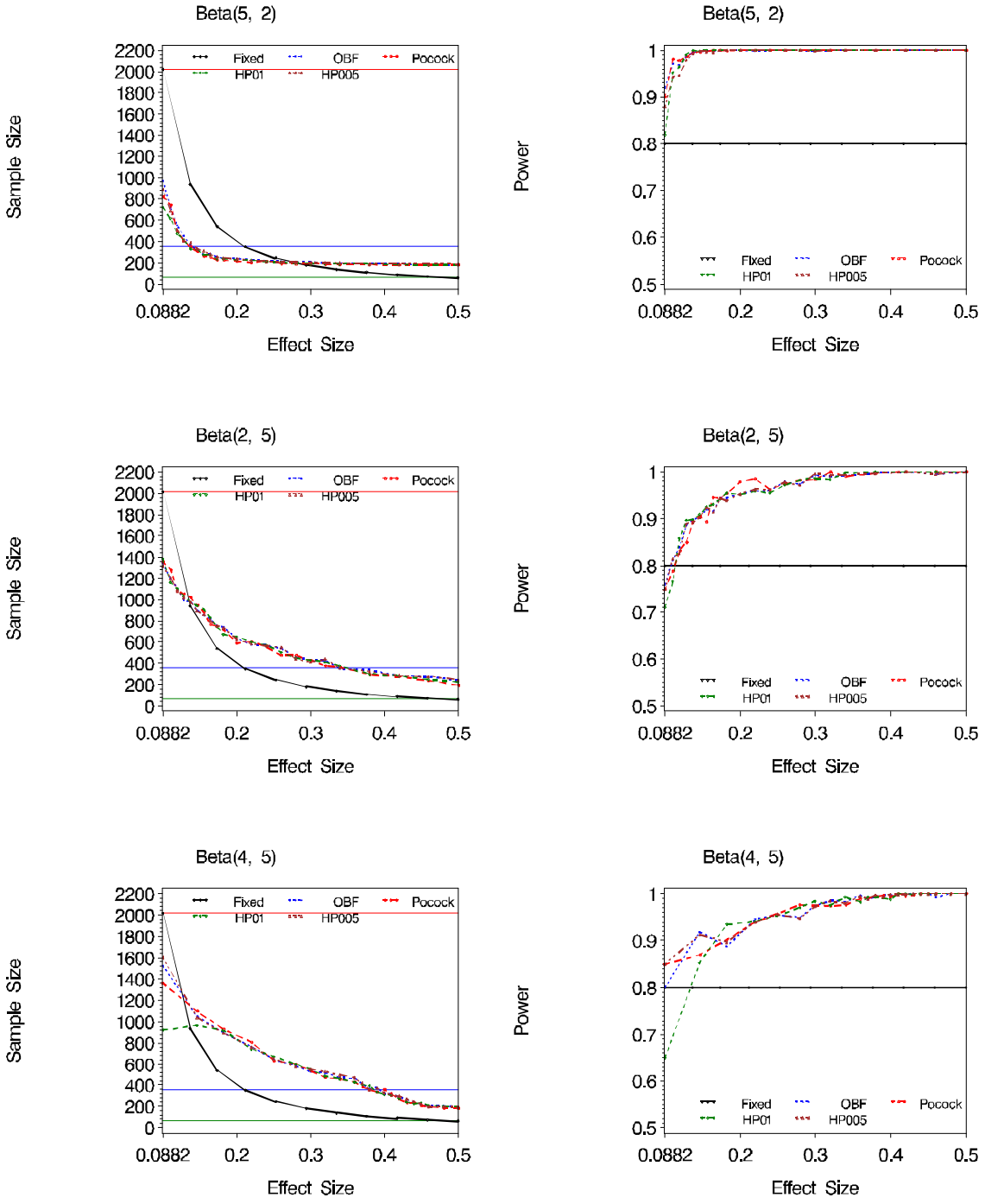

Figure C.7: Sample Size and Power Curves for Two-look Unweighted Sample Size Re-estimation Designs when Treatment Effects follow Beta Distributions

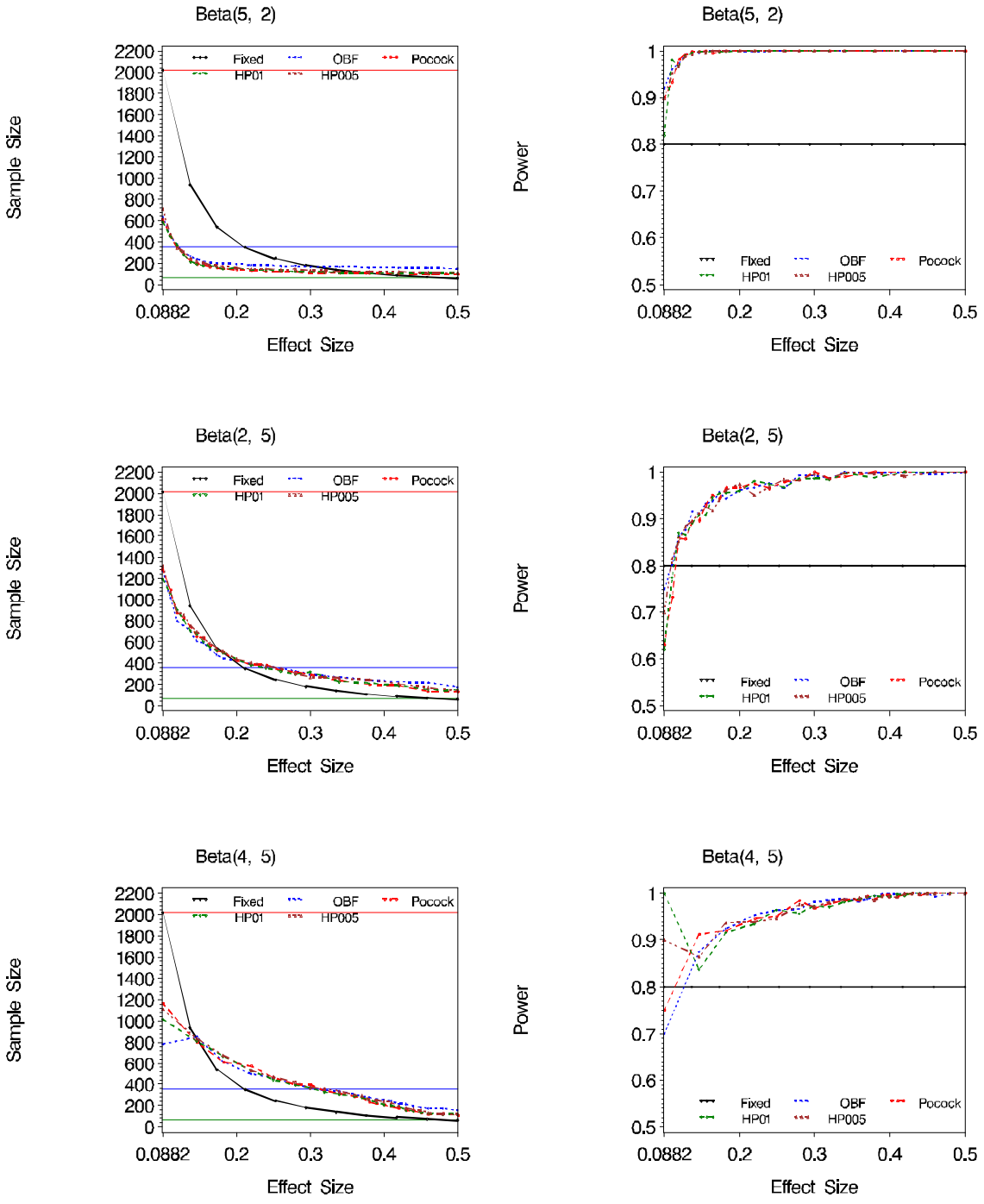

Figure C.8: Sample Size and Power Curves for Five-look Equal-spaced Unweighted Sample Size Re-estimation Designs when Treatment Effects follow Beta Distributions (re-estimation at look 3)

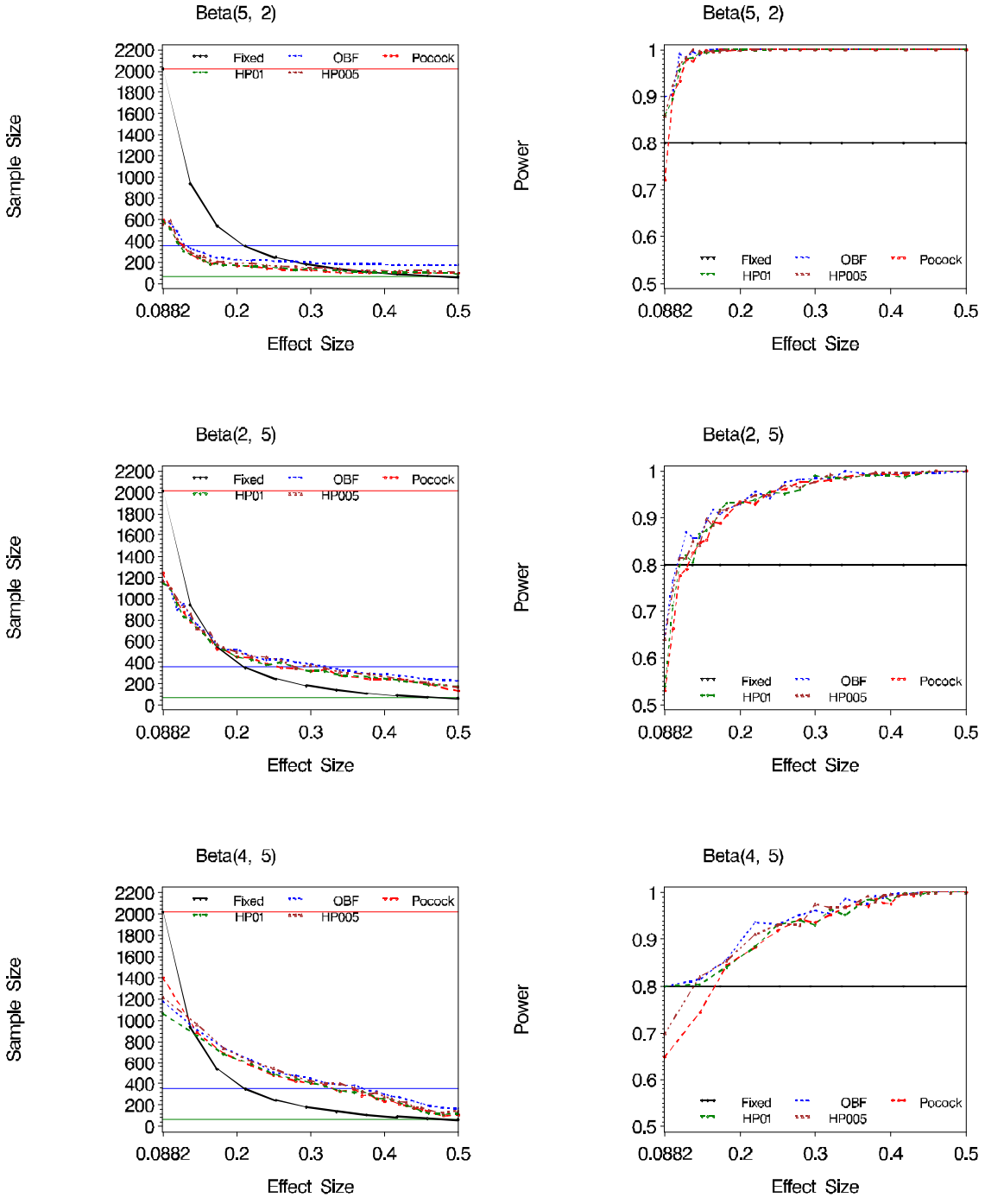

Figure C.9: Sample Size and Power Curves for Five-look Unequal-spaced Unweighted Sample Size Re-estimation Designs when Treatment Effects follow Beta Distributions (re-estimation at look 3)

Bibliography

[1] Aravantinos G, Fountzilas G, Kosmidis P et al. Paclitaxel plus carboplatin versus paclitaxel plus alternating carboplatin and cisplatin for initial treatment of advanced ovarian cancer: long-term efficacy results: a Hellenic Cooperative Oncology Group (HeCOG) study. *Annals of Oncology*, 16:1116-1122, 2005

[2] Armitage P, McPherson K, and Rowe BC. Repeated significance tests on accumulating data *Journal of the Royal Statistical Society*, 132:235-244, 1969

[3] Chow YS and Teicher H Probability Theory: Independence, Interchangability, Martingales, 3rd edition *Springer*, 2003

[4] Chuang-Stein C Seizue the opportunities. *Pharmaceutical Statistics*, 3:157-159, 2004

[5] Cui L, Hung HM, Wang SJ. Modification of Sample Size in Group Sequential Clinical Trials. *Biometrics*, 55:853-857, 1999

[6] Du Bois A, Luck HJ, Meier W et al. A Randomized Clinical Trial of Cisplatin/Paclitaxel Versus Carboplatin/Paclitaxel as First-Line Treatment of Ovarian Cancer. *Journal of the National Cancer institute*, 95:1320-1330, 2003

[7] Dunnett C. A multiple comparison procedure for comparing several treatments with a control. *Journal of the American Statistical Association*, 50:1096-1121, 1955

[8] East 5.1: A software package for the design and interim monitoring group sequential clinical trials *Cambridge, MA: Cytel Software Corporation*, 2007

[9] Fisher B, Costantino JP, Wickerham DL et al. Tamoxifen for prevention of breast cancer: report of the National Surgical Adjuvant Breast and Bowel Project P-1 Study. *Journal of National Cancer Institute*, 90:1371-1388, 1998

[10] Genz, A. Computing Numerical computation of multivariate normal probabilities. *Journal of Computational and Graphical Statistics*, 1:141-150, 1992

[11] Gould, AL and Shih, WJ. Sample size re-estimation without unblinding for normally distributed outcomes with unknown variance. *Communication in Statistics, Theory and Methods*, 21: 2833-2853, 1992

[12] Gould, AL and Shih, WJ. Modifying the design of ongoing trials without unblinding. *Statistics in Medicine*, 17:89-100, 1998

[13] Guidance for industry: Adaptive design clinical trials for drugs and biologics, 2010

[14] Hybittle, J.L. repeated assessment of results in clinical trials of cancer treatment. *Journal of Radiology*, 44:793-797, 1971

[15] Jennison C and Turnbull BW. Statistical approaches to interim monitoring of medical trials: A review and commentary. *Statistical Science*, 5:299-317, 1990

[16] Jennison C and Turnbull BW. Effcient group sequential designs when there are several effect sizes under consideration. *Statistics in Medicine*, 25:917-932, 2006

[17] Jung SH, Kim C, and Chow SH. Sample size calculation for the log-rank tests for multi-arm trials with a control. *Journal of the Korean Statistical Society*, 37:11-22, 2008

[18] Lan KKG and DeMets D. Interim analysis: The alpha spending function approach. *Statistics in Medicine*, 13:1341-1352, 1994

[19] Lan KKG, Soo Y, and Shun Z. Normal approximation for two-stage winner design. In Random Walks, Sequential Analysis and Related Topics, Hsiung AC, Ying Z, Zhang CH (eds). *World Scientific: Singapore*, 28-43, 2006

[20] Lesko LJ. Tools to reduce phase III trial failures. *Proceedings of the AGAH annual meeting*, 2006

[21] Li G, Shih WJ, Xie T, and Lu J. A sample size adjustment procedure for clinical trials based on conditional power. *Biostatistics*, 3:277–287, 2002

[22] Li G, Shih WJ, and Wang Y. Two-stage adaptive design for clinical trials with survival data. *Journal of Biopharmaceutical Statistics*, 15:707-718, 2005

[23] Liu GF, Zhu R, and Cui L. Evaluating the adaptive performance of flexible sample size designs with treatment difference in an interval. *Statistics in Medicine*, 27:584-596, 2008

[24] Marshall AW and Olkin I. A multivariate exponential distribution. *Journal of the American Statistical Association*, 62:30-44, 1967

[25] Neijt JP, Engelholm SA, Tuxen MK et al. Exploratory phase III study of paclitaxel and cicplatin versus paclitaxel and carboplatin in advanced ovarian cancer. *Journal of Oncology*, 18:3084-3092, 2000

[26] O'Brien PC and Fleming TR. A multiple testing procedure for clinical trials. *Biometrics*, 35:549-556, 1979

[27] Ozols RF, Bundy BN, Greer BE et al. Phase III Trial of Carboplatin and Paclitaxel Compared With Cisplatin and Paclitaxel in Patients With Optimally Resected Stage III Ovarian Cancer: A Gynecologic Oncology Group Study. *Journal of Clinical Onlcology*, 21:3194-3200, 2003

[28] Parmar MKB, Adams M, Balestrino M et al. Paclitaxel plus carboplatin versus standard chemotherapy with either single-agent carboplatin or cyclophosphamide, doxorubicin, and cisplatin in women with ovarian cancer: the ICON3 randomised trial. *The Lancet*, 360:505-515, 2002

[29] Peto, R., Pike, M.C., Armitage, P., Breslow, N.E., Cox, D.R., Howard, D.R., Mantel, N., McPherson, K., Peto, J., and Smith, P.G. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. I. Introduction and design. *British Journal of Cancer*, 34:585-612, 1976

[30] Pocock, SJ. Group sequential methods in the design and analysis of clinical trials. *Biometrika*, 64:191-199, 1977

[31] Proschan MA, Lan KKG, and Wittes JT. Statistical monitoring of clinical trials. *Springer, New York*, 2006

[32] Royston P, Parmar MKB, and Qian W. Novel designs for multi-arm clinical trials with survival outcomes with an application in ovarian cancer. *Statistics in Medicine*, 22:2239-2256, 2003

[33] Schaid DJ, Wieand S, and Therneau TM Optimal two-stage screening designs for survival comparisons. *Biometrika*, 77:507-513, 1990

[34] Shih WJ Group sequential, sample size re-estimation and two-stage adaptive designs in clinical trials: a comparison. *Statistics in Medicine*, 25:933-941, 2006

[35] Shun Z, Lan KKG, and Soo Y Interim treatment selection using the normal approximation approach in clinical trials. *Statistics in Medicine*, 27:597-618, 2008

[36] Thoelke KR. Creating a Framework for Success: A Clinical Strategy for FDA Approval. *Proceedings of the Conference: From Pipeline to Product: Navigating the FDA Approval Process*, 2007.

[37] Tsiatis AA. The asymptotic joint distribution of the efficient scores test for the proportional hazards model calculated over time. *Biometrika*, 68:311-315, 1981

[38] Tsiatis AA. Repeated significance testing for a general class of statistics used in censored survival analysis. *Journal of the American Statistical Association*, 77:855-861, 1982

[39] Tsiatis AA and Mehta C On the inefficiency of the adaptive design for monitoring clinical trials. *Biometrika*, 90:367-378, 2003

[40] Veronesi U, Maisonneuve P, Sacchini V et al. Tamoxifen for breast cancer among hysterectomised women. *Lancet*, 359:1122-1124, 2002

[41] Wittes J and Brittain E. The role of pilot studies in increasing the efficiency of clinical trials (with comments). *Statistics in Medicine*, 9:65-72, 1990

Vita

Fang Fang

EDUCATION

1995 B.S. in Nursing, Tianjin Medical University, Tianjin, China.
2001 M.S. in Biostatistics, University of Illinois, Chicago, IL.
2010 Ph.D. in Biostatistics, University of Medicine & Dentistry of New Jersey.

EXPERIENCE

01/01-08/01 Statistical Analyst (consultant), Astellas Pharma, Inc (former Fujisawa Healthcare), Deerfield, IL
08/01-12/03 Statistician, Astellas Pharma, Inc, Deerfield, IL
12/03-09/06 Manager, Eisai Medical Research, Inc., Ridgefield Park, NJ
09/06-04/09 Senior Manager, Eisai Medical Research, Inc., Ridgefield Park, NJ
04/09-05/09 Associate Director, Eisai Medical Research, Inc., Ridgefield Park, NJ
06/09-10/10 Associate Director, Bayer Healthcare Pharmaceuticals, Inc. Montville, NJ