

**USE OF SIMULTANEOUS INFERENCE UNDER  
ORDER RESTRICTION, STEPDOWN TESTING  
PROCEDURE AND STAGE-WISE SEQUENTIAL  
OPTIMAL DESIGN IN CLINICAL DOSE STUDY**

by

**Gang Jia**

B.S. Management Science, Fudan University, 1991

M.A. Statistics, University of Pittsburgh, 1998

Submitted to the Graduate Faculty of  
the Department of Statistics in partial fulfillment  
of the requirements for the degree of

**Doctor of Philosophy**

University of Pittsburgh

2004

UNIVERSITY OF PITTSBURGH

DEPARTMENT OF STATISTICS

This dissertation was presented

by

Gang Jia

It was defended on

December 6th, 2004

and approved by

Leon J. Gleser, Department of Statistics

Allan R. Sampson, Department of Statistics

Henry W. Block, Department of Statistics

Harry S. Wieand, Department of Biostatistics

Dissertation Director: Leon J. Gleser, Department of Statistics

**USE OF SIMULTANEOUS INFERENCE UNDER ORDER RESTRICTION,  
STEPDOWN TESTING PROCEDURE AND STAGE-WISE SEQUENTIAL  
OPTIMAL DESIGN IN CLINICAL DOSE STUDY**

Gang Jia, PhD

University of Pittsburgh, 2004

This dissertation discusses the design approaches of adaptive dose escalation study and the analysis methods of dose study data, and the relationship between the study design approach and data analysis methods.

A general max-min approach to construct simultaneous confidence intervals for the monotone means of correlated and normally distributed random samples is proposed to analyze correlated dose response data. The approach provides an accurate, flexible and computationally easy way to obtain critical values of simultaneous confidence intervals under monotone order restriction.

A stepdown testing procedure for analyzing dose study data is examined and a modified stepdown testing approach is proposed to incorporate the adaptive sampling nature of the study data. A mixture normal approximate approach of the dose response is proposed to analyze the binary outcome with small sample size at the first stage of the adaptive design.

Finally, an optimal stage-wise adaptive clinical dose study design is proposed to be applied in a dose escalation study with binary outcome and correlated dose response. The

study design criteria is defined as a weighted average power to identify all effective dose levels. A back-induction algorithm is used to obtain the optimal design parameters. The values of optimal design parameters vary when different analysis methods are used to analyze the study data.

Simulation studies are performed to illustrate the two proposed analysis methods and the proposed optimal design approach.

## TABLE OF CONTENTS

<b>PREFACE</b> . . . . .	x
<b>1.0 INTRODUCTION AND BACKGROUND</b> . . . . .	1
1.1 Introduction . . . . .	1
1.2 Background . . . . .	2
1.3 Research Outline . . . . .	4
<b>2.0 SIMULTANEOUS CONFIDENCE INTERVALS OF MONOTONE DOSE RESPONSE</b> . . . . .	7
2.1 Introduction . . . . .	7
2.2 Problem and Mathematic Model . . . . .	8
2.3 Literature Review . . . . .	9
2.4 Korn's Max-Min Procedure . . . . .	11
2.5 Simultaneous Confidence Intervals for Multiple Comparison Under Restriction of Isotonic Dose Response . . . . .	12
2.5.0.1 Proof of the Extension of Korn's Procedure in Correlated dose response variables: . . . . .	13
2.5.1 Critical Values of Simultaneous Confidence Intervals for Correlated Dose Response Variables . . . . .	15
2.6 Application of The General Max-min Approach to Binary Dose Response Data	18

2.6.1	Measurements of Drug Effect for Binary Dose Response . . . . .	18
2.6.2	Simulation Study . . . . .	19
<b>3.0</b>	<b>STEPDOWN TESTING PROCEDURES FOR DOSE FINDING STUDY WITH ADAPTIVE DESIGN . . . . .</b>	<b>30</b>
3.1	Introduction . . . . .	30
3.2	Literature Review on Stepwise Testing Procedure in Dose Study . . . . .	31
3.3	Closed Testing Procedure . . . . .	33
3.4	Modified Stepdown Testing Procedures in Adaptive Sampling . . . . .	35
3.5	A Stepdown Testing Procedure with Adaptive Sampling Inference . . . . .	38
3.6	Application of the Modified Stepdown Testing Procedure in Adaptive Dose Study with a Binary Endpoint . . . . .	39
3.6.1	Expected conditional distribution approach for stepdown testing in clinical study with binary outcome . . . . .	40
3.6.2	A Monte Carlo Simulation Approach to Obtain the Critical Values . . . . .	43
3.6.2.1	Example . . . . .	44
<b>4.0</b>	<b>OPTIMAL STAGE-WISE ADAPTIVE CLINICAL DOSE STUDY DESIGN . . . . .</b>	<b>49</b>
4.1	Introduction . . . . .	49
4.2	Literature Review on Adaptive Phase II Clinical Study Designs . . . . .	50
4.3	An Adaptive Study Design for Dose Escalation Study . . . . .	53
4.3.1	Proposed Study Design . . . . .	53
4.3.2	Optimal Adaptive Design When Stepdown Testing Procedure is Used to Analyze the Study Data . . . . .	58
4.3.2.1	A Conservative Stage-Wise Optimization Approach Under Least Favorable Configurations . . . . .	60

4.3.2.2 A Back-Induction Algorithm to Solve for the Optimal Cut Points	63
4.3.3 Optimal Adaptive Design When Simultaneous Confidence Intervals are Used to Analyze the Study Data . . . . .	67
4.3.4 Case Study . . . . .	70
<b>5.0 FUTURE RESEARCH PLAN . . . . .</b>	<b>76</b>
<b>APPENDIX A. SAS PROGRAMS USED TO DERIVE THE CRITICAL VALUES OF MODIFIED SIMULTANEOUS CONFIDENCE INTER- VALS . . . . .</b>	<b>78</b>
<b>APPENDIX B. SAS PROGRAMS USED TO DERIVED THE CRITICAL VALUES FOR MIXTURE NORMAL APPROXIMATION APPROACH</b>	<b>93</b>
<b>APPENDIX C. SAS PROGRAMS USED TO OBTAIN THE OPTIMAL CUT POINTS WHEN SCI APPROACHED IS USED . . . . .</b>	<b>100</b>
<b>APPENDIX D. SAS PROGRAMS USED TO OBTAIN THE OPTIMAL CUT POINTS WHEN STEPDOWN TESTING APPROACHED IS USED . . . . .</b>	<b>109</b>
<b>BIBLIOGRAPHY . . . . .</b>	<b>117</b>

## LIST OF TABLES

1	Critical Values for Simultaneous Confidence Intervals . . . . .	22
2	Coverage Probability(%) of 95% Simultaneous Confidence Intervals of Correlated Dose Response . . . . .	23
3	Coverage Probability(%) of 95% Simultaneous Confidence Intervals Using Dunnett's Critical value . . . . .	24
4	Coverage Probability(%) of 95% Simultaneous Confidence Intervals Using Hayter's Critical Value . . . . .	25
5	Coverage Probability(%) of 95% Simultaneous Confidence Intervals Using Tukey's Critical Value . . . . .	26
6	Coverage Probability(%) of 95% Simultaneous Confidence Intervals of Correlated Dose Response with Smaller Sample Size . . . . .	27
7	Coverage Probability(%) of 95% Simultaneous Confidence Intervals of Correlated Dose Response with 50% Smaller Expected Event Rates . . . . .	28
8	Adjusted Critical Values for One-sided Test with $\alpha=0.1$ . . . . .	45
9	Adjusted Critical Values for One-sided Test with $\alpha=0.05$ . . . . .	46
10	Adjusted Critical Values for One-sided Test with $\alpha=0.025$ . . . . .	47
11	Adaptive Dose Study Design: 200 Subjects Enrolled at the First Stage of Each Dose Level . . . . .	73



12	Adaptive Dose Study Design: 300 Subjects Enrolled at the First Stage of Each Dose Level . . . . .	74
13	Adaptive Dose Study Design: 400 Subjects Enrolled at the First Stage of Each Dose Level . . . . .	75

## PREFACE

I would like to express my deepest gratitude to my advisor, Professor Leon J. Gleser, for his understanding, patience, guidance and encouragement. Without his support, this work would not have been done. Dr. Gleser has always had his door and his warm heart open whenever I need help, both professionally and emotionally.

I would also like to thank the members of my thesis committee, Professor Harry S. Wieand, Professor Henry W. Block and Professor Allan R. Sampson for their generous help and valuable inputs. And I would like to thank the Department of Statistics for giving me the opportunity to finish my doctoral study here.

Last but certainly not least, I would like to thank my wife, Lingfang, and my parents, Zesheng Jia and Yunqin Wang, for their continuous support throughout this journey. I would not have been able to complete it without them. I would like to thank my son, Alan, and my baby girl, Katherine, for their understanding. This work is dedicated to all of you.

## 1.0 INTRODUCTION AND BACKGROUND

### 1.1 INTRODUCTION

Dose study is an important part in drug development. A safe and potentially effective dose strategy needs to be identified before a full-scale phase III clinical trial can start. Especially in areas involving severe or life threatening outcomes, due to the usually very low event rate which in turn requires large sample size, only one or two dose groups are feasible in the final study.

Phase II clinical trials are conducted to screen experimental drugs and select promising candidates to be tested in Phase III confirmatory trials. Information obtained in Phase II studies also provides critical information for the efficient design of Phase III trials. Traditionally, Phase II trials have focused on a single dose level that has been determined in Phase I studies. MED-Minimum Effective Dose, is often used when toxicity is of primary concern, and MTD-Maximum Tolerated Dose, is often used when the targeted disease is severe and the risk of not being sufficiently treated is high. In modern drug development, however, it is increasingly common for potential therapies to enter into Phase II testing where the best choice of dose level is less clear. This is often the case in modern cancer therapeutics and cardiovascular medicines, where many new agents under development work in biologically complex ways and might not have a steep dose-response relationship.

## 1.2 BACKGROUND

The proposed research has been motivated by a multi-center Phase II clinical trial that was coordinated in the Cleveland Clinic Foundation Heart Center. I participated in the trial as a project statistician. The trial was conducted to assess the effect of a new cardiovascular drug to reduce the risk of procedure-related restenosis and subsequent adverse cardiac events. A brief background of the experimental drug and its targeted disease are described in the following paragraph from the study protocol:

Restenosis is a gradual loss of vessel lumen diameter resulting from thrombogenic and wound healing processes at the site of vessel wall injury following endovascular interventions and graft surgery. Restenosis is the predominant cause of failure following PTCA and hemodialysis shunt insertions, occurring at rates of 30% and 50% respectively. Morbidity and secondary interventions due to restenosis following PTCA (Percutaneous Transluminal Coronary Angioplasty / Stent) is estimated to cost 2,500 lives and 2.5 billion dollars per year in the U.S. No medical therapy is available for effective prevention of restenosis. Heparin is the commonly used drug in patients undergoing PTCA or other invasive vascular procedures. While heparin reduces the risk of acute thrombosis, it has no beneficial effect on restenosis. TTD, a new cardiovascular medicine, was being developed to provide safe medical prevention of immediate thrombus formation and restenosis. TTD exerts its beneficial effect by diminishing the triggering of coagulation and is therefore expected to be more efficient than other anti-thrombotic drugs in preventing the generation of thrombin and recruitment of platelets.

The objective of this trial was to compare the efficacy of 6 different dosing strategies of TTD with heparin in patients undergoing elective or urgent PTCA with respect to treating the incidence of death, Myocardial Infarction and/or urgent revascularization of the target coronary vessel within the index hospitalization or 7 days following the index PTCA. The safety of different doses of TTD in patients undergoing elective or urgent PTCA was also assessed with respect to the incidence of major and minor bleeding, death and hemorrhagic stroke. The result of the Phase II trial was used to assess whether there exists a potentially superior effect of the new drug indicating that a Phase III confirmatory trial should be con-

ducted to verify that it can be marketed as an alternative treatment. The logical endpoint in a cardiovascular medicine study is usually mortality, or a composite endpoint of recurrent Myocardial Infarction, urgent target vessel revascularization and/or mortality. An alternative endpoint that can be used is the treatment success of the cardiovascular drug used with the PTCA procedure, which can be defined as the significant reduction of lumen stenosis. Recent development of the IVUS technique, an approach that utilizes ultrasound to measure the lumen wall and vessel volume, make it possible to precisely measure the actual thickness of harmful plaque inside the coronary vessel. The significant reduction of the stenosis is an important indicator of the long term success of the treatment.

Due to the severe nature of the disease, the new drug had to be administrated in conjunction with a partial dose of heparin. On the other hand, the potential side effects of the new drug are also very serious, including major bleeding, disabling stroke and deadly intracranial hemorrhage. Hence it was decided by the medical staff that a conservative dose escalation approach would be used. The study would start from the lowest dose level. The results of the lower dose groups would be analyzed before escalating to the next higher dose. By escalating the dose level of the new drug, one hopes to see an increasing drug effect in treating the target cardiovascular disease. The dose escalation continues until all of the planned dose levels are studied, or unacceptable increased risk of the adverse events is observed. Zero-dose placebo is not ethically acceptable here, so an active control arm with full-dose heparin is used for comparison. It is assumed that the effect size increases as the dose of new drug increases, and so does the risk of adverse events. Thus it is natural and efficient to incorporate the directional assumption in the statistical design and analysis of the study. Because the level of maximum tolerance has been studied in the Phase I study, stopping the dose escalation due to safety concerns is not expected to happen and the study will most likely to go through all the planned dose levels. Therefore, adjustment for overall

Type I error for safety monitoring won't be necessary.

### 1.3 RESEARCH OUTLINE

The thesis can be outlined as two major parts: the adaptive experiment design, and the analysis of the dose study data. At the experiment design stage, we propose an adaptive design at each dose level to improve the efficiency of subject allocation. Starting from the lowest dose level, the data will be analyzed after part of the planned number of subjects are enrolled at the dose level. If the efficacious effect is not obvious enough, we will move to the next higher dose level without enrolling the rest of the subjects and will subsequently assign saved subjects to the remaining higher dose levels. By doing that, we will have more power to analyze the drug effect if it only exists at higher dose levels, and can have a better estimate of the effect size. On the other hand, although the study drug is presumably more effective at high dose levels, we still need to have enough subjects at lower levels to study its possible adverse effects, and to detect its potential effect at lower levels. Indeed, estimating the effective dose range is also an important part of the study objective. A wider effective dose range means that the drug can be used in broader patient population. After the study enrollment is finished and follow up data collected, we propose to use the approaches of simultaneous confidence intervals and stepdown testing procedure to analyze the study data.

Topics on sequential adaptive design have been researched extensively, especially in the area of clinical trials. However, in the area of clinical dose studies that have multiple dose groups with monotone increasing / decreasing drug response, there has not been adequate research on sequential adaptive design to improve study designs using the order restriction

information. It is often assumed that the conventional hypothesis testing and/or confidence region approaches will be used at the end of the study. As the first objective of the thesis, I propose an optimal adaptive study design that is tailored based on the analysis approach to be used to analyze the study data. A flexible study design criteria is defined using the weighted expected proportion of identified effective dose levels.

At the data analysis stage, single-step or stepwise multiple testing procedures will be explored to compare the control group and each of the dose levels. The parametric regression strategy would be normally considered to analyze the dose-response relationship. However, the most important objection to such an approach is that we are usually interested in estimations at the extreme dose range, where the choice of the form of regression model is much more critical. Moreover, the experiment itself, rarely consisting of more than 6-7 dose levels in a Phase II study, will provide insufficient evidence for determining the correct regression model.

We need to perform the data analysis to find those dose levels which are effective and to estimate their effect sizes relative to the control; such information will be important in making the decision whether to conduct the Phase III trial and will also provide crucial information for the design of Phase III trial. A stepdown testing procedure (Chapter 3) is a powerful approach to identify the effective levels, as the type I error probability of each individual test doesn't need to be penalized to ensure that the overall family-wise type I error probability is controlled. On the other hand, the approach of simultaneous confidence intervals has the advantage over stepdown testing procedures in that it provides interval estimates of the effect size, and can be used to verify a hypothesized parametric dose response relationship. Simultaneous confidence intervals are usually overly conservative. As the second part of this thesis, ways to improve the simultaneous confidence intervals, such as only looking at effective dose levels and incorporating the assumption of monotone effective sizes into the

construction of the confidence intervals, will be studied. The third part of the thesis will be to explore the stepwise testing procedures when the data is from studies with sequential adaptive designs.

The data to be collected in this study is dichotomous, recorded as treatment success or whether an endpoint event occurred. The therapeutic effect of the new drug relative to the active controlled arm in such cases is usually measured using odds ratio, relative risk, or difference of event rates. A problem arises at the early stages of the trial when the number of subjects enrolled in the controlled arm is low, and the number of events that occurs is not big enough for any asymptotic-theory-based inference to be used. Thus, it will be necessary to explore and evaluate the performance of various normal based approaches when they are applied in binomial settings. We will address this practical issue in each of three parts of the thesis.

Because our proposed adaptive design involves the analysis approaches that are used to analyze the study data, we organize the thesis in the following order: Chapter 2 discusses simultaneous confidence intervals for correlated drug effect estimates under order restriction, Chapter 3 discusses stepdown testing procedures for ordered dose effect with adaptive sampling data, and Chapter 4 discusses a stage-wise sequential adaptive design in Phase II clinical dose studies. The last chapter discusses my plans for future research in the three areas studies in this thesis.



## 2.0 SIMULTANEOUS CONFIDENCE INTERVALS OF MONOTONE DOSE RESPONSE

### 2.1 INTRODUCTION

In a clinical dose study, it is often of interest to have confidence bands for dose-response curves. These confidence bands can be used to form simultaneous confidence intervals for drug response at multiple doses, or to form a confidence interval for doses yielding a given drug response. If a parametric family of dose-response relationships can be assumed, the parameters and the curve can be estimated using a regression approach, and the simultaneous confidence intervals can be obtained based on the estimates of these parameters. However, in studies of new drug development, more often than not such a parametric assumption is not available, except that one can reasonably assume that the underlying dose response is non-increasing or non-decreasing depending on the nature of the drug effect.

A natural estimator of the dose response under such an isotonic assumption is the isotonic regression, which is the conventional least squares estimate of the dose response curve under the isotonic restriction. The approach of simultaneous confidence intervals we will study in this chapter is based on the same idea of isotonic regression, aimed to utilize the assumption of isotonic dose response to achieve accurate and efficient interval estimates.

## 2.2 PROBLEM AND MATHEMATIC MODEL

In a Phase II clinical dose study, drug response data  $y_{i1}, y_{i2}, \dots, y_{in_i}$  have been collected at dose levels  $d_i, i = 0, 1, 2, \dots, k$ . Dose level  $d_0$  represents the control arm, which can be a zero-dose placebo group or an active control group treated with a different drug. The  $y_{ij}$  are assumed to be normally distributed with mean  $\mu_i$  and variance  $\sigma_i^2$ . Although in most occasions, the variance is not known, we will assume that it is known here, and the extension of our approach to the cases with unknown variance is straightforward. We assume that the mean drug responses at dose levels  $d_1, d_2, \dots, d_k$  are non-increasing:  $\mu_1 \geq \mu_2 \geq \dots \geq \mu_k$ . Notice that the drug response of zero-dose arm  $\mu_0$  is not included in the inequality.

The drug effect at dose level  $d_i$  is defined as  $\theta_i = \mu_i - \mu_0$ , the relative therapeutic effect of the experiment drug compared to the placebo or active control. It is obvious that the dose effects defined in this way also satisfy the inequality  $\theta_1 \geq \theta_2 \geq \dots \geq \theta_k$ . The simultaneous confidence intervals with  $1 - \alpha$  coverage probability are defined as

$$P\{\theta_i \in (l_i, u_i) : i = 1, 2, \dots, k\} = 1 - \alpha \quad (2.1)$$

Notice that estimates of  $\theta_i, 1 \leq i \leq k$ , will be correlated because all are measured relative to the same control. In this chapter, we will introduce an approach to derive the simultaneous confidence intervals based on the above relative dose effect variables using a max-min procedure and the maximum modulus of the multivariate distribution. The max-min procedure was first suggested by Korn [14](1982) to construct simultaneous confidence intervals for independent dose response variables under isotonic restriction. We will prove that this approach can be extended to correlated dose response variables.

It is well known that Tukey's procedure based on the Studentized range distribution and Scheffé's procedure based on the  $F$ -distribution enable us to make multiple comparisons

among a set of sample means while controlling the overall simultaneous probability coverage. However, in practice we often only need to make comparisons between one common control and each of the experimental treatment groups. In our case, as well as in many other practical applications, we also have prior knowledge about the order of the expected means of treatment groups. Although procedures by Tukey and Scheffé still apply, they are too conservative and it is possible to derive more precise confidence intervals.

### 2.3 LITERATURE REVIEW

One of the earliest attempts to increase the efficiency of Scheffé's confidence bounds for multiple comparisons by incorporating additional information was by Bohrer in 1967. Bohrer [3] suggested in his paper a sharper extension of Scheffé's confidence bounds for linear functions of independent normal random variables when assuming all coefficients of the linear function are nonnegative. Bohrer gave tables of the critical values, and also showed that when the sample size is large, the common width of the improved simultaneous confidence bounds is about 70.7% of the common width of Scheffé's confidence bounds.

Marcus et al [19](1976) developed flexible simultaneous lower confidence bounds for the means in a one-way ANOVA model when the means are restricted to a certain subset of  $R_k$  as long as this subset is closed under multiplication by a positive scalar. They gave the result when all means are nonnegative and the result when the means are ordered. Marcus [17](1982) pointed out that the approach can also be used to obtain two-sided simultaneous confidence intervals.

D. A. Williams [36](1977) discussed the limiting distributions of the estimated maximum and range of a set of ordered normal means when all means are in fact equal. Williams showed

that approximate  $1 - \alpha$  simultaneous confidence intervals for all ordered contrasts of the normal means can be obtained by using the percentile points of these limiting distributions.

Dunnett [8] [9](1955 and 1964) developed an approach to obtain simultaneous confidence intervals for comparing the means of multiple treatment groups to the mean of a common control group when the response variable is normally distributed with all groups having a common variance. Tables of exact critical values for one-sided and two-sided confidence limits are given when the assumed common variance  $\sigma^2$  is known or when it is not known. In problems when the outcome is a proportion, which is the case in a lot of clinical studies, a variable transformation  $y_j = \arcsin p_j^{1/2}$  of the same proportions  $p$  can be used. When the sample sizes  $n_j$  within each group are reasonably large,  $y_j$  can be treated as normally distributed with variance  $\sigma^2 = \frac{1}{4}$ .

Korn [14](1982) proposed a simple procedure using a modification of the Studentized maximum modulus technique to construct confidence bands for dose-responses under nondecreasing order restriction. This procedure assumes no parametric model for the dose-response curves.

Schoenfeld [22] (1986) suggested using the intersection of a  $1 - \alpha$  confidence region for the mean response vector  $(\mu_1, \mu_2, \dots, \mu_k)$  and the set of all monotone non-decreasing or non-increasing functions on dose levels to construct  $1 - \alpha$  simultaneous confidence intervals for  $\mu_i$ .

Hayter [12](1990) proposed to use a one-sided studentized range test (OSRT) to construct a simultaneous lower confidence bound for all ordered pair-wise contrasts. C. C. Lee [15](1996) improved Korn's procedure by introducing the distribution of a generalization of the maximum modulus.

Our review of methodology research in the literature tells us that although extensive work

has been done in the area of simultaneous confidence intervals under the restriction of isotonic dose response, results have been almost exclusively limited to the assumption of independent dose response variables. However, as I will argue later, this assumption often fails to hold in applications. Dose response variables are usually correlated. Thus developing simultaneous confidence intervals under order restrictions for correlated dose response variables can have practical importance. In the next section I will give an detailed discussion of Korn's max-min procedure because it is closely related to the approach I will propose in this chapter.

## 2.4 KORN'S MAX-MIN PROCEDURE

Korn's [14](1982) procedure using a modification of the Studentized maximum modulus technique provides a simple, intuitive, yet efficient way to construct simultaneous confidence intervals under order restriction. This procedure assumes no parametric model for the dose-response curves. Here, we will briefly introduce this procedure and will also provide the proof of Korn's claim under more general conditions.

Suppose dose response data  $y_{i1}, y_{i2}, \dots, y_{in_i}$  are collected at dose level  $x_i, i = 1, \dots, k$ . The  $y_{ij}$  are assumed to be normally distributed with mean  $f(x_i)$  and common unknown variance  $\sigma^2$ , and write  $\bar{y}_i$  as the sample mean of dose group  $i$ . Model-free simultaneous  $1 - \alpha$  confidence intervals are given by

$$\bar{y}_i - m_{k,n-k}s/\sqrt{n_i} \leq f(x_i) \leq \bar{y}_i + m_{k,n-k}s/\sqrt{n_i} \quad (2.2)$$

for  $i = 1, 2, \dots, k$ , where  $s^2$  is the usual pooled estimate of the variance  $\sigma^2$ , and  $m_{k,n-k}$  is the upper  $\alpha$  point of the Studentized maximum modulus distribution with parameters  $k, n - k$ . If the dose-response curve  $f(x)$  is known to be non-decreasing, then the above confidence

intervals can be modified as

$$\max_{l \leq i} \{\bar{y}_l - m_{k,n-k} s / \sqrt{n_l}\} \leq f(x_i) \leq \min_{j \geq i} \{\bar{y}_j + m_{k,n-k} s / \sqrt{n_j}\} \quad (2.3)$$

for  $i = 1, 2, \dots, k$ . These max-min confidence intervals are derived from the sample means and the isotonic assumption on  $f(x)$ : the  $f(x_i)$  satisfy (2.2) and are a non-decreasing sequence if and only if the  $f(x_i)$  satisfy (2.3). Thus, the simultaneous confidence intervals (2.3) have exactly  $1 - \alpha$  coverage probability. It is possible that these intervals do not contain the usual isotonic regression estimates, and may even be empty. However, the coverage probability is true if the monotone assumption is correct. Korn also pointed out at the end of his paper that the same technique can be used in deriving simultaneous confidence intervals for contrasts between all doses and a common control using Dunnett's approach, under the assumption that the dose response relationship is nondecreasing. However, he didn't give further details or examples for the extension.

## 2.5 SIMULTANEOUS CONFIDENCE INTERVALS FOR MULTIPLE COMPARISON UNDER RESTRICTION OF ISOTONIC DOSE RESPONSE

Clinical dose studies in most therapeutic areas are comparative in nature. Subjectiveness of drug responses, heterogeneity of patient population, uncontrollable effect of different medical practice and continuing development in concomitant medications are among the many reasons that the absolute drug responses are not very good indication of the actual drug effect. In these cases, a control group, either a placebo group that receives only a dummy drug, or an active control group that receives a different treatment (usually some treatment

commonly used in standard practice), are enrolled in parallel with the experimental drug group. Randomization and blinding is often used to ensure unbiased and random patient recruitment. In a comparative dose study, the approach of comparing each dose group to the common control group, as introduced earlier in this chapter, is the most common analysis method used.

The following proof shows that the max-min procedure by Korn can be extended to the cases of correlated isotonic dose response variables:

**2.5.0.1 Proof of the Extension of Korn's Procedure in Correlated dose response variables:** Suppose  $y_{i1}, y_{i2}, \dots, y_{in_i}$  are random normal variables distributed as  $N(\mu_i, \sigma)$ ,  $i = 1, 2, \dots, k$ , and  $\bar{y}_i$  is the sample mean. Suppose  $q_k^\alpha$  is the critical value for  $1 - \alpha$  simultaneous confidence interval such that

$$\Pr(\underline{\mu} \subset Q) = 1 - \alpha \quad (2.4)$$

where

$$Q = \{\bar{y}_i - q_k^\alpha \sigma / \sqrt{n_i} \leq \mu_i \leq \bar{y}_i + q_k^\alpha \sigma / \sqrt{n_i}, i = 1, 2, \dots, k\} \quad (2.5)$$

Now additionally suppose that we know the order restriction that  $(\mu_1, \mu_2, \dots, \mu_k) \in O$ , where  $O = (\mu_1 \leq \mu_2 \leq \dots, \leq \mu_k) \cap R^k$ , and we define the confidence set  $\tilde{Q}$  as:

$$\tilde{Q} = \left\{ \begin{array}{l} \max_{l \leq i} \{\bar{y}_l - q_k^\alpha \sigma / \sqrt{n_l}\} \leq \mu_i \leq \min_{j \geq i} \{\bar{y}_j + q_k^\alpha \sigma / \sqrt{n_j}\}, \\ i = 1, 2, \dots, k \end{array} \right\}$$

It is obvious that  $\tilde{Q} \subseteq Q$ . Therefore

$$\begin{aligned}
& Pr\{(\mu_1, \mu_2, \dots, \mu_k) \in \tilde{Q}\} \\
& \leq Pr\{(\mu_1, \mu_2, \dots, \mu_k) \in Q\} \\
& = 1 - \alpha.
\end{aligned}$$

On the other hand, for any point  $A = (a_1, a_2, \dots, a_k) \in O \cap Q$ :

$$\begin{aligned}
A \in O & \Rightarrow a_1 \leq a_2 \leq \dots \leq a_k \\
A \in Q & \Rightarrow \bar{y}_i - q_k^\alpha \sigma / \sqrt{n_i} \leq a_i \leq \bar{y}_i + q_k^\alpha \sigma / \sqrt{n_i}, i = 1, 2, \dots, k
\end{aligned}$$

So the point  $A$  has to satisfy

$$\begin{aligned}
\max_{l \leq i} \{\bar{y}_l - q_k^\alpha \sigma / \sqrt{n_l}\} & \leq a_i \leq \min_{j \geq i} \{\bar{y}_j + q_k^\alpha \sigma / \sqrt{n_j}\}, i = 1, 2, \dots, k \\
\Rightarrow A & \in \tilde{Q}
\end{aligned}$$

In other words,  $Q \cap O \subseteq \tilde{Q}$ . Since we already know  $(\mu_1, \mu_2, \dots, \mu_k) \in O$ , the following is also true:

$$\begin{aligned}
& Pr\{(\mu_1, \mu_2, \dots, \mu_k) \in \tilde{Q}\} \\
& \geq Pr\{(\mu_1, \mu_2, \dots, \mu_k) \in Q \cap O\} \\
& = Pr\{(\mu_1, \mu_2, \dots, \mu_k) \in Q\} \\
& = 1 - \alpha
\end{aligned}$$

Therefore we have proven that the simultaneous confidence interval  $\tilde{Q}$  is the exact confidence set:

$$Pr\{(\mu_1, \mu_2, \dots, \mu_k) \in \tilde{Q}\} = 1 - \alpha$$



under the order restriction  $\mu_1 \leq \mu_2 \leq \dots, \leq \mu_k$ . Notice that the proof doesn't require that the  $y_{ij}$ 's have to be independent, nor does it require that the variances have to be the same.

In fact, the max-min procedure can also be expressed as

$$\begin{aligned} \tilde{Q} = \{ & \inf\{\mu_i : (\mu_1, \mu_2, \dots, \mu_k) \in Q \cap O\} \\ & \leq \mu_i \leq \\ & \sup\{\mu_i : (\mu_1, \mu_2, \dots, \mu_k) \in Q \cap O\}, i = 1, 2, \dots, k\}. \end{aligned}$$

The proof is valid for any convex subset  $O$  of the parameter space, where a convex subset guarantees that the resulting simultaneous confidence intervals are dense. Therefore the above inequality can be applied to other similar distributions such as the multivariate t-distribution when the  $\sigma_i^2$ 's are not known. Korn [14] studied the efficiency achieved by using the max-min procedure for independent normal distributions. Korn's simulation result show that when the means are in fact equal, the max-min procedure can reduce the average width of the simultaneous confidence intervals of six normal means by approximately 30%, and the reduction increases to approximately 50% for ten normal means.

### 2.5.1 Critical Values of Simultaneous Confidence Intervals for Correlated Dose Response Variables

Dunnett's [8] [9] results provided the critical values for comparing multiple treatment groups to a common control. But Dunnett's approach only applies to the cases where all the observed treatment groups' responses have the same variance  $\sigma_t^2$ , although it allows a different variance  $\sigma_c^2$  in the common control group. This translates into the simple equicorrelated covariance structure. Even in cases when it is usually reasonable to assume equal variance of drug response among treatment groups, with the increasing use of adaptive study designs, we often have different sample sizes among the treatment groups. Thus the estimated drug

response variables have different variances, and the resulting covariance matrix often has complicated and unpredictable structure.

Somerville [24][25](1997 and 2001) developed computation methodology to calculate the critical values for a large class of simultaneous confidence intervals. Given  $k$  populations with unknown means  $\mu_i$ ,  $i = 1, 2, \dots, k$ , assume that the location parameter estimates  $x_i$  have a joint multivariate normal distribution with mean  $\mu$  and non-singular covariance matrix  $\sigma^2\Sigma$  with  $\Sigma$  known and  $s^2$  an independent estimate of  $\sigma^2$  with  $\nu$  degrees of freedom. Suppose  $B = \{c_1, c_2, \dots, c_m\}$  is a finite collection of linear functions of the  $x_i$ 's. Somerville shows that, his approach can solve for the critical value  $q$  such that

$$Pr\left(c'_i x / (\text{var}(c'_i x))^{1/2} \leq q/\sqrt{2} : c_i \in B, i = 1, 2, \dots, m\right) = 1 - \alpha \quad (2.6)$$

for arbitrary values for  $c_i, m, k, \nu, \Sigma$  within acceptable computing time on computers with Pentium III processor. Critical values for a large class of simultaneous intervals can be obtained using this approach, including Tukey's all pairwise comparisons, Dunnett's one-sided and two-sided comparisons with a control, Williams-type comparisons, Marcus-type comparisons, Hayter's one-sided simultaneous confidence intervals and others.

Using a computer program based on Somerville's computing algorithm, critical values for simultaneous confidence intervals of correlated dose responses variables can be easily obtained by specifying appropriate covariance structure and contrast vectors. In a typical dose response study as described at the beginning of the Chapter,  $k$  dose groups and a common control group are used, and the dose response in each group is assumed to be normally distributed with means  $\mu_0, \mu_i, i = 1, 2, \dots, k$  and variance  $\sigma_i^2$ . The drug effect at each dose group is measured as  $\theta_i = \mu_i - \mu_0, i = 1, 2, \dots, k$ . Suppose the study design allocates sample size  $n_i$  to group  $i, i = 1, 2, \dots, k$ . The observed drug effects at each dose  $\bar{x}_i = \bar{y}_i - \bar{y}_0, i = 1, 2, \dots, k$ , have a multivariate normal distribution with mean  $(\theta_1, \theta_2, \dots, \theta_k)$ ,

and covariance matrix  $\Sigma$ , whose diagonal elements are  $\sigma_{ii}^2 = \sigma_i^2/n_i + \sigma_0^2/n_0$ , and off-diagonal elements are  $\sigma_{ij}^2 = \sigma_0^2/n_0$  ( $i \neq j$ ). Although the covariances are all the same between each pair of dose effect estimates, the correlation coefficients are different unless all the dose groups have the same sample size. The relative dose effect can also be viewed as applying the contrast matrix  $(c_1, c_2, \dots, c_k)$  to dose response  $(\mu_0, \mu_1, \dots, \mu_k)$ , where each  $c_i$  is a contrast vector with first element equal to  $-1$ ,  $i$ th element equal to  $1$ , and all the rest are zeros.

The computing algorithm can be used to calculate the boundary values for the  $1 - \alpha$  simultaneous confidence intervals of  $k$  dose effect estimates, say  $q_{\rho,k}^\alpha$ , such that

$$P\left\{ \begin{array}{l} \bar{y}_i - \bar{y}_0 - q_{\rho,k}^\alpha(\sigma_i\sqrt{1/n_i} + \sigma_0\sqrt{1/n_0}) \leq \theta_i \leq \bar{y}_i - \bar{y}_0 + q_{\rho,k}^\alpha(\sigma_i\sqrt{1/n_i} + \sigma_0\sqrt{1/n_0}), \\ i = 1, 2, \dots, k \end{array} \right\} = 1 - \alpha \quad (2.7)$$

We assume the isotonic restriction on dose effect such that  $\theta_1 \geq \theta_2 \geq \dots \geq \theta_k$ . By applying the max-min procedure, we have the following simultaneous confidence intervals under isotonic dose response restriction

$$\begin{aligned} & \max_{l \leq i} \{ \bar{y}_l - \bar{y}_0 - q_{\rho,k}^\alpha(\sigma_l\sqrt{1/n_l} + \sigma_0\sqrt{1/n_0}) \} \\ & \leq \mu_i - \mu_0 \leq \\ & \min_{j \geq i} \{ \bar{y}_j - \bar{y}_0 - q_{\rho,k}^\alpha(\sigma_j\sqrt{1/n_j} + \sigma_0\sqrt{1/n_0}) \}, \quad i = 1, 2, \dots, k \end{aligned} \quad (2.8)$$

The computing algorithm is Monte-Carlo simulation based and can achieve extremely high accuracies of critical values by increasing the Monte-Carlo sampling sizes. Therefore, under correct assumption of multivariate normal distributions, the above simultaneous confidence intervals can be treated as having the exact coverage probability  $1 - \alpha$ .

## 2.6 APPLICATION OF THE GENERAL MAX-MIN APPROACH TO BINARY DOSE RESPONSE DATA

In this chapter we will apply the approach of simultaneous confidence intervals in a clinical trial with the endpoint defined as a binary variable. In the dose study we introduced in Chapter 1, the binary outcome can be defined as the occurrence of death, MI and/or urgent revascularization of the target coronary vessel within certain period after cardiovascular drug treatment, or the treatment success defined as successful reduction of vessel stenosis.

### 2.6.1 Measurements of Drug Effect for Binary Dose Response

Let  $p_i$  be the observed event rate at dose  $i$  and let  $p_0$  be the event rate for the control group, the drug effect at each dose relative to control treatment can be measured using one of the following four measurements:

- Difference of observed event rate between each dose level and the control:  $(p_i - p_0)$ .
- Difference of arcsine of the square root of event rate  $(\arcsin p_i^{\frac{1}{2}} - \arcsin p_0^{\frac{1}{2}})$ .
- The log odds ratio between each dose level and the control  $(\ln \frac{p_i(1-p_0)}{(1-p_i)p_0})$ .
- Log of Relative risk between each dose level and the control  $(\ln p_i/p_0)$ .

Although Brown et al [5] showed that the coverage probability of normal-approximation-based confidence intervals for binomial data can be unstable for even moderate sample size, he also pointed out that relative measures such as difference of proportions perform much better. Therefore, we propose to use the normal distribution based approach we have developed to obtain simultaneous confidence intervals for the above four comparisons. The covariance matrix for each of the four measurements can be easily identified with the knowledge of sample sizes in the treatment groups. Suppose the sample size in dose group  $i$  is  $n_i$ , the

corresponding number of observed adverse events is  $m_i$ , and the estimated event rate  $\hat{p}_i = m_i/n_i$ ,  $i = 0, 1, 2, \dots, k$ . The estimated covariances of the approximate multivariate normal distributions of the above four measurements of the estimated relative dose effects at dose 1 to  $k$  are:

- $(p_i - p_0)$ :  $\sigma_{ii} = \frac{\hat{p}_i(1-\hat{p}_i)}{n_i} + \frac{\hat{p}_0(1-\hat{p}_0)}{n_0}$ ,  $\sigma_{ij(i \neq j)} = \frac{\hat{p}_0(1-\hat{p}_0)}{n_0}$
- $(\arcsin p_i^{\frac{1}{2}} - \arcsin p_0^{\frac{1}{2}})$ :  $\sigma_{ii} = \frac{1}{2}$ ,  $\sigma_{ij(i \neq j)} = \frac{1}{4}$
- $(\ln \frac{p_i(1-p_0)}{(1-p_i)p_0})$ :  $\sigma_{ii} = \frac{1}{n_i - m_i} + \frac{1}{m_i} + \frac{1}{n_0 - m_0} + \frac{1}{m_0}$ ,  $\sigma_{ij(i \neq j)} = \frac{1}{n_0 - m_0} + \frac{1}{m_0}$
- $(\ln p_i/p_0)$ :  $\sigma_{ii} = \frac{1-\hat{p}_i}{m_i} + \frac{1-\hat{p}_0}{m_0}$ ,  $\sigma_{ij(i \neq j)} = \frac{1-\hat{p}_0}{m_0}$

## 2.6.2 Simulation Study

The above four different measures of drug effect are explored. Continuity adjustment is done by adding 0.5 to the event count. Ten different monotone dose response configurations for six dose levels are used, and 5000 Monte Carlo simulations are used in each case. The true event rates for each dose response configuration are as follows:

### Dose Response

Configuration	Control	Dose 1	Dose 2	Dose 3	Dose 4	Dose 5	Dose 6
1	0.10	0.10	0.10	0.10	0.10	0.10	0.10
2	0.10	0.12	0.12	0.12	0.12	0.12	0.12
3	0.10	0.08	0.08	0.08	0.08	0.08	0.08
4	0.10	0.12	0.11	0.10	0.09	0.08	0.07
5	0.10	0.12	0.12	0.10	0.10	0.08	0.08
6	0.10	0.12	0.12	0.12	0.08	0.08	0.08
7	0.10	0.09	0.08	0.07	0.06	0.05	0.04
8	0.10	0.17	0.16	0.15	0.14	0.13	0.12
9	0.10	0.12	0.08	0.08	0.08	0.08	0.08
10	0.10	0.12	0.12	0.12	0.12	0.12	0.08

These dose response configurations include flat dose response (1 2 3), linear dose response (4 7 8) and step dose response (5 6 9 10). Table 1 compares the critical values of simultaneous confidence intervals for correlated multivariate normal means obtained by Somerville's simulation algorithm and those provided by Dunnett [9], Tukey and Hayter. Covariance matrices used were based on the dose configurations and drug response measurements. For measurements using relative risk  $p_i/p_0$  or difference of risk  $p_i - p_0$ , the covariance matrix also depends on the sample estimates of event rate. However, we calculated the critical values based on the known true probabilities for the purpose of comparison. When using simulation to validate the coverage probabilities (Table 2-7), we calculate the critical values based on the sample probabilities from each random sample. It can be seen that the critical values we have calculated in most cases are fairly close to Dunnett's approach. However, the confidence intervals based on the exact critical values when the actual covariance structure is taken into

consideration have generally more accurate coverage probability (Table 2). Given easy and relatively fast computation and the wide availability of high-speed micro computers, the exact critical values definitely are preferred.

Simulation results of the coverage probability of 95% simultaneous confidence intervals are presented in Table 2 to Table 7. The results in Tables 2 to 5 are based on sample sizes of 1500 in the control group, 500 in the first two or four dose groups, and 1000 in the last four or two dose groups. These sample sizes reflect a simple adaptive design that assigns more subjects to higher dose groups and the control group to increase overall efficiency. The simultaneous confidence intervals obtained using the max-min procedure and critical values based on correlated normal means perform the best (Table 2). The actual coverage probabilities are very close to 95% even with moderate sample size, and are consistent across different dose response configurations. The results from using Dunnett's critical values are not far off (Table 3), but are clearly not as accurate. On the other hand, we can also see that the simultaneous confidence intervals using Hayter's approach (Table 4) and Tukey's approach (Table 5) are too conservative, usually having actual coverage probability from 0.965 to 0.985.

It is well known that the normal approximation approach for binary variables usually does not perform well when the sample sizes are small, or when the expected binomial proportions are close to 0 or 1. Table 2.6 lists the simulation results when the sample sizes for each group are reduced by 90%, i.e., 150 in the control group, 50 in the first two or four dose groups, and 100 in the last four or two dose groups. It is comforting to see that the performance is actually fairly good. Table 2.7 lists the simulation results when the binomial proportions for control and each dose group are cut by 50%. The control group has expected event rate at 5%, and so on. When the drug effect is measured as differences between event rates or differences between arcsin square root of event rates, the coverage probabilities are still close

Table 1: Critical Values for Simultaneous Confidence Intervals

Dose Config.	$p_i - p_0$	$\arcsin p_i^{\frac{1}{2}} - \arcsin p_0^{\frac{1}{2}}$	$\ln \frac{p_i(1-p_0)}{(1-p_i)p_0}$	$\ln p_i/p_0$
Sample Size: Control=1500 Doses 1-2=500 Dose 3-6=1000				
1	2.60232	2.60225	2.60232	2.56773
2	2.60828	2.60225	2.59280	2.55457
3	2.59310	2.60225	2.61098	2.58045
4	2.59894	2.60225	2.60523	2.57180
5	2.60113	2.60225	2.60337	2.56769
6	2.60038	2.60225	2.60392	2.56849
7	2.58011	2.60225	2.61753	2.59227
8	2.61238	2.60225	2.58213	2.54110
9	2.59607	2.60225	2.60867	2.57656
10	2.59604	2.60225	2.60766	2.57698
Sample Size: Control=1500 Doses 1-4=500 Dose 5-6=1000				
1	2.60996	2.60990	2.60996	2.56773
2	2.61459	2.60990	2.60308	2.55457
3	2.60332	2.60990	2.61662	2.58045
4	2.60734	2.60990	2.61267	2.57180
5	2.60920	2.60990	2.61130	2.56769
6	2.60909	2.60990	2.61086	2.56849
7	2.59297	2.60990	2.62135	2.59227
8	2.61749	2.60990	2.59450	2.54110
9	2.60538	2.60990	2.61427	2.57656
10	2.60594	2.60990	2.61420	2.57698
Dunnett=2.61611 Hayter=2.82734 Tukey=3.02916				



Table 2: Coverage Probability(%) of 95% Simultaneous Confidence Intervals of Correlated Dose Response

Dose Config.	$p_i - p_0$	$\arcsin p_i^{\frac{1}{2}} - \arcsin p_0^{\frac{1}{2}}$	$\ln \frac{p_i(1-p_0)}{(1-p_i)p_0}$	$\ln p_i/p_0$
Sample Size: Control=1500 Doses 1-2=500 Dose 3-6=1000				
1	0.94666	0.94758	0.94614	0.95166
2	0.94650	0.94778	0.94438	0.95142
3	0.94508	0.94628	0.94544	0.95134
4	0.94576	0.94686	0.94554	0.95140
5	0.94600	0.94700	0.94446	0.95094
6	0.94644	0.94860	0.94648	0.95250
7	0.94586	0.94610	0.94780	0.95196
8	0.94834	0.94918	0.94424	0.95270
9	0.94648	0.94772	0.94706	0.95216
10	0.94438	0.94758	0.94700	0.95308
Sample Size: Control=1500 Doses 1-4=500 Dose 5-6=1000				
1	0.94306	0.94518	0.94500	0.95092
2	0.94362	0.94608	0.94366	0.95088
3	0.94012	0.94328	0.94406	0.95052
4	0.94272	0.94464	0.94466	0.95040
5	0.94228	0.94436	0.94310	0.94928
6	0.94350	0.94624	0.94542	0.95158
7	0.94090	0.94338	0.94700	0.95096
8	0.94678	0.94782	0.94252	0.95140
9	0.94276	0.94572	0.94606	0.95158
10	0.94096	0.94508	0.94530	0.95142

Table 3: Coverage Probability(%) of 95% Simultaneous Confidence Intervals Using Dunnett's Critical value

Dose Config.	$p_i - p_0$	$\arcsin p_i^{\frac{1}{2}} - \arcsin p_0^{\frac{1}{2}}$	$\ln \frac{p_i(1-p_0)}{(1-p_i)p_0}$	$\ln p_i/p_0$
Sample Size: Control=1500 Doses 1-2=500 Dose 3-6=1000				
1	0.94086	0.94212	0.94616	0.94668
2	0.94020	0.94292	0.94628	0.94746
3	0.94114	0.94028	0.94452	0.94442
4	0.94106	0.94158	0.94536	0.94604
5	0.94036	0.94116	0.94458	0.94564
6	0.94164	0.94316	0.94650	0.94720
7	0.94336	0.94048	0.94468	0.94448
8	0.94154	0.94378	0.94910	0.95026
9	0.94180	0.94184	0.94632	0.94648
10	0.93946	0.94150	0.94600	0.94718
Sample Size: Control=1500 Doses 1-4=500 Dose 5-6=1000				
1	0.93710	0.93972	0.94502	0.94528
2	0.93696	0.94102	0.94536	0.94684
3	0.93622	0.93706	0.94346	0.94292
4	0.93768	0.93990	0.94436	0.94472
5	0.93660	0.93864	0.94318	0.94394
6	0.93874	0.94050	0.94552	0.94570
7	0.93824	0.93758	0.94398	0.94388
8	0.94002	0.94220	0.94712	0.94840
9	0.93840	0.94002	0.94560	0.94544
10	0.93666	0.93928	0.94442	0.94574

Table 4: Coverage Probability(%) of 95% Simultaneous Confidence Intervals Using Hayter's Critical Valus

Dose Config.	$p_i - p_0$	$\arcsin p_i^{\frac{1}{2}} - \arcsin p_0^{\frac{1}{2}}$	$\ln \frac{p_i(1-p_0)}{(1-p_i)p_0}$	$\ln p_i/p_0$
Sample Size: Control=1500 Doses 1-2=500 Dose 3-6=1000				
1	0.96158	0.96384	0.96670	0.96710
2	0.96206	0.96362	0.96728	0.96754
3	0.96178	0.96148	0.96554	0.96532
4	0.96196	0.96254	0.96536	0.96538
5	0.96122	0.96268	0.96612	0.96626
6	0.96252	0.96336	0.96672	0.96672
7	0.96366	0.96184	0.96568	0.96526
8	0.96358	0.96604	0.96930	0.97030
9	0.96242	0.96258	0.96616	0.96632
10	0.96190	0.96394	0.96718	0.96742
Sample Size: Control=1500 Doses 1-4=500 Dose 5-6=1000				
1	0.95886	0.96236	0.96648	0.96680
2	0.95930	0.96208	0.96636	0.96618
3	0.95792	0.95976	0.96564	0.96502
4	0.95884	0.96096	0.96532	0.96538
5	0.95836	0.96084	0.96558	0.96552
6	0.95960	0.96220	0.96616	0.96594
7	0.96020	0.95944	0.96572	0.96494
8	0.96178	0.96524	0.96866	0.96934
9	0.95910	0.96092	0.96604	0.96582
10	0.95856	0.96216	0.96662	0.96656

Table 5: Coverage Probability(%) of 95% Simultaneous Confidence Intervals Using Tukey's Critical Value

Dose Config.	$p_i - p_0$	$\arcsin p_i^{\frac{1}{2}} - \arcsin p_0^{\frac{1}{2}}$	$\ln \frac{p_i(1-p_0)}{(1-p_i)p_0}$	$\ln p_i/p_0$
Sample Size: Control=1500 Doses 1-2=500 Dose 3-6=1000				
1	0.97934	0.98020	0.98306	0.98306
2	0.97842	0.98042	0.98242	0.98284
3	0.97878	0.97954	0.98200	0.98248
4	0.97896	0.97974	0.98200	0.98212
5	0.97866	0.97944	0.98186	0.98214
6	0.97952	0.98014	0.98194	0.98242
7	0.97922	0.97836	0.98148	0.98158
8	0.98078	0.98216	0.98450	0.98476
9	0.97900	0.97940	0.98194	0.98236
10	0.97916	0.98044	0.98226	0.98240
Sample Size: Control=1500 Doses 1-4=500 Dose 5-6=1000				
1	0.97754	0.97964	0.98290	0.98294
2	0.97636	0.97970	0.98206	0.98258
3	0.97646	0.97872	0.98170	0.98226
4	0.97696	0.97884	0.98198	0.98192
5	0.97642	0.97834	0.98164	0.98172
6	0.97714	0.97922	0.98178	0.98234
7	0.97640	0.97744	0.98146	0.98148
8	0.97946	0.98180	0.98366	0.98398
9	0.97664	0.97858	0.98214	0.98254
10	0.97704	0.97960	0.98194	0.98232

Table 6: Coverage Probability(%) of 95% Simultaneous Confidence Intervals of Correlated Dose Response with Smaller Sample Size

Dose Config.	$p_i - p_0$	$\arcsin p_i^{\frac{1}{2}} - \arcsin p_0^{\frac{1}{2}}$	$\ln \frac{p_i(1-p_0)}{(1-p_i)p_0}$	$\ln p_i/p_0$
Sample Size: Control=150 Doses 1-2=50 Dose 3-6=100				
1	0.94602	0.94998	0.96940	0.97368
2	0.94414	0.94546	0.96612	0.97304
3	0.95320	0.94590	0.96578	0.96824
4	0.94860	0.95152	0.96990	0.97366
5	0.94630	0.94812	0.96752	0.97142
6	0.94872	0.94682	0.96900	0.97272
7	0.95380	0.94932	0.96250	0.96142
8	0.93914	0.94754	0.96482	0.97448
9	0.95008	0.94676	0.96750	0.97008
10	0.94392	0.94690	0.96950	0.97384
Sample Size: Control=150 Doses 1-4=50 Dose 5-6=100				
1	0.94042	0.94764	0.96688	0.96916
2	0.93640	0.94502	0.96498	0.96858
3	0.94662	0.94450	0.96200	0.96020
4	0.94226	0.94882	0.96562	0.96752
5	0.93980	0.94472	0.96550	0.96700
6	0.93988	0.94564	0.96608	0.96684
7	0.94948	0.94994	0.95558	0.95266
8	0.93206	0.94572	0.96536	0.97294
9	0.94340	0.94504	0.96376	0.96256
10	0.93704	0.94662	0.96706	0.96966

Table 7: Coverage Probability(%) of 95% Simultaneous Confidence Intervals of Correlated Dose Response with 50% Smaller Expected Event Rates

Dose Config.	$p_i - p_0$	$\arcsin p_i^{\frac{1}{2}} - \arcsin p_0^{\frac{1}{2}}$	$\ln \frac{p_i(1-p_0)}{(1-p_i)p_0}$	$\ln p_i/p_0$
Sample Size: Control=1500 Doses 1-2=500 Dose 3-6=1000				
1	0.94208	0.94464	0.94764	0.95352
2	0.94308	0.94540	0.94574	0.95330
3	0.93968	0.94150	0.94864	0.95408
4	0.94150	0.94410	0.94786	0.95252
5	0.94222	0.94436	0.94700	0.95266
6	0.94338	0.94526	0.94818	0.95374
7	0.94134	0.94008	0.95018	0.95434
8	0.94420	0.94768	0.94638	0.95420
9	0.94118	0.94402	0.94950	0.95476
10	0.93966	0.94512	0.94960	0.95426
Sample Size: Control=1500 Doses 1-4=500 Dose 5-6=1000				
1	0.93618	0.94184	0.94684	0.95340
2	0.93876	0.94298	0.94636	0.95228
3	0.93314	0.93736	0.94854	0.95340
4	0.93646	0.94176	0.94708	0.95236
5	0.93662	0.94142	0.94568	0.95174
6	0.93808	0.94224	0.94786	0.95330
7	0.93486	0.93700	0.95088	0.95482
8	0.93758	0.94374	0.94414	0.95138
9	0.93402	0.94066	0.94878	0.95452
10	0.93314	0.94206	0.94892	0.95318

to the target of 95%. But the simultaneous confidence intervals are rather conservative when odds ratios or relative risks are used. In cases where one of these two measures is desired, methods using the exact binomial distributions can be used to obtain the boundary points of confidence intervals at all dose levels, and similar max-min procedures can be applied to derive the simultaneous confidence intervals of the isotonic dose responses.

### 3.0 STEPDOWN TESTING PROCEDURES FOR DOSE FINDING STUDY WITH ADAPTIVE DESIGN

#### 3.1 INTRODUCTION

The problem of identifying the lowest effective drug dose for which the mean response is superior to that at the control group is often an important issue in a drug efficacy study. Likewise, one of the most important goals in a toxicity study is to identify the highest safe dose, which is defined as the highest safe drug dose for which the mean safety response is not worse than that at the control group. Stepdown testing procedure is a widely used approach in this area. Under the assumption of a monotone mean configuration, the stepdown testing procedure is particularly powerful and easy to perform. In this chapter we will concentrate on the dose studies for finding the lowest effective dose. The testing approach for finding the highest safe dose can be formulated in exactly the same way.

Suppose the drug response variables have means  $\mu_i$  at dose levels  $d_i$ ,  $i = 1, 2, \dots, k$ , and  $\mu_0$  is the mean drug response of the control group. Drug efficacy is measured as the relative drug response such as  $\mu_i - \mu_0$ . Write the means of relative drug efficacy as  $\theta_i$ ,  $i = 1, 2, \dots, k$ . Dose level  $d_i$  is defined to be effective if  $\theta_i > 0$ . Notice that the dose responses don't have to be independent of each other, as we have discussed in the previous chapter. Suppose that we can assume  $\theta_1 \leq \theta_2 \leq \dots \leq \theta_k$ . Then  $d_E$  is the lowest effective dose if  $\theta_j \leq 0$  for  $j < E$  and  $\theta_j > 0$  for  $E \leq j \leq k$ . In practice, people usually define the lowest effective



dose in a dose study as dose level  $d_E$  if  $\theta_j \leq 0$  for  $j < E$  and  $\theta_j \geq \delta$  ( $\delta > 0$ ) for  $E \leq j \leq k$ .  $\delta$  here is usually referred to as the smallest (clinically) meaningful drug effect. The use of  $\delta$  is useful in comparing statistical approaches, because no statistically procedure can discriminate between the effective and noneffective dose groups if the means of the dose effects can be arbitrarily close to each other [31].

The design and validity of a stepdown testing procedure to find the lowest effective dose level follows a general procedure given by Marcus, Peritz and Gabriel [18]. For a family of hypotheses  $\{H_{0i} : i = 1, 2, \dots, k\}$  in a multiple testing problem, the family-wise error rate(FWE) is defined as

$$\text{FWE} = Pr\{\text{at least one true } H_{0i} \text{ is rejected}\}.$$

A step testing procedure controls the family-wise error rate strongly if FWE is controlled when any subgroup of the family of null hypotheses are true. If a step testing procedure satisfies the conditions of closed procedure, which we will define in section 3.3, the procedure controls the family-wise error rate strongly.

### 3.2 LITERATURE REVIEW ON STEPWISE TESTING PROCEDURE IN DOSE STUDY

Stepwise testing procedures in dose study with isotonic dose responses have been extensively researched. The earliest well known result is by Williams [34] [35] (1971 and 1972). Williams proposed a stepwise testing procedure to compare the mean responses of  $k$  dose levels with a zero-dose control by using the isotonic regression estimates. The table of upper  $\alpha$  percentage points of the distribution of the comparisons is provided by Williams for equal allocation of subjects (1971) and for unequal allocation of subjects (1972).

Tamhane et al [28](1996) did a comparison study of the stepwise testing procedures using William's approach and various linear contrasts. He showed that William's procedure, the stepdown procedure using Helmert Contrast and critical value of correlated  $t$ -statistics, and the stepdown procedure using linear contrast and critical value for simple  $t$ -statistics have the highest average powers, and their performances are relatively stable for different values of MED (Minimum Effective Dose),  $\delta$  (smallest meaningful difference between treatment and control) and the type of dose response function.

Bauer [1] (1997) in a follow-up paper pointed out that the strong control of family-wise Type I error rate is guaranteed in stepwise procedures only if monotone order restriction is assumed to hold among response means. The claim made by Tamhane [28] et al (1996) that the monotone assumption in response means is not required is only true for many-to-one pairwise comparisons.

Bauer et al [2] (2001) proposed stepwise procedures when both efficacy and safety need to be considered. A stepwise procedure for suitably defined subfamilies is used to control the overall family-wise error, and order restrictions for both efficacy and safety are assumed. Similarly, Tamhane et al [29] (2002) provided a stepwise testing procedure for detecting the minimum effective (MINED) and maximum safe dose (MAXSD) using the joint  $t$ -distribution of efficacy response and safety response variables at each dose level.

Channon et al [6] (2001) pointed out that although stepdown testing procedures provide powerful tools for identifying the minimum effective and/or maximum safe dose, simultaneous inference about dose responses after the testing is still desirable in practice. Channon et al performed simulation studies to compare the actual coverage of the proposed simultaneous confidence intervals. It is shown that although the minimum coverage is not theoretically guaranteed, the actual coverage ranges from 93% to 97%, which should be satisfactory for

practical purposes.

Bretz et al [4] (2003) suggested that usually the relative difference between treatments and control is of interest instead of absolute measures of treatments. They proposed to conduct stepdown procedures based on the ratio of the drug responses between each dose level and the control.

One common shortcoming of the previous research is that the proposed testing procedures have implicitly assumed that the dose study data are randomly collected from the probability distributions. In a clinical study with adaptive design, this assumption is not correct because the data collected are not totally random samples. It is necessary to tailor stepwise testing procedures for non-random samples in order to draw valid statistical inference. We will try to address this problem in this chapter.

### 3.3 CLOSED TESTING PROCEDURE

Let  $\{H_{0i} : 1 \leq i \leq k\}$  be a finite family of null hypotheses. One can form the closure of this family by taking all the nonempty intersections  $H_{0p} = \bigcap_{i \in P} H_{0i}$  for  $\emptyset \subset P \subseteq \{1, 2, \dots, k\}$ . If an  $\alpha$ -level test of each hypothesis  $H_{0p}$  is available, then a closed testing procedure can be constructed by rejecting any null hypothesis  $H_{0p}$  only if every  $H_{0q}$  is rejected by its associated  $\alpha$ -level test for all  $H_{0q} \supseteq H_{0p}$ . As pointed out by Marcus et al [18](1976), a closed testing procedure strongly controls the Type I family-wise error probability in multiple comparisons, i.e., the Type I family-wise error probability is controlled under any configuration of the null hypotheses family.

Suppose in a dose study, one can assume that the drug responses at control group and dose levels 1 to  $k$  are normal random variables with the means as  $\mu_i, i = 0, 1, 2, \dots, k$ , and

a common variance  $\sigma^2$ . One can further assume that  $\mu_1 \leq \mu_2 \leq \dots \leq \mu_k$ . Suppose that the drug effect at each dose level is defined as the difference between the drug response of the control and each dose group:  $\theta_i = \mu_i - \mu_0, i = 1, 2, \dots, k$ . We can estimate the mean drug effects by sample mean differences  $\bar{y}_i - \bar{y}_0, i = 1, 2, \dots, k$ , with corresponding standard errors as  $(\frac{1}{n_0} + \frac{1}{n_i})^{\frac{1}{2}}\sigma$ , where  $n_0, n_1, \dots, n_k$  are the sample sizes in each group correspondingly. Consider the family of null hypotheses  $H_{0i} : 0 \geq \theta_i \geq \theta_{i-1} \geq \dots \geq \theta_1, i = 1, 2, \dots, k$ , and the corresponding alternative hypotheses  $H_{1i} : \theta_i > 0$ . It is obvious that, under the isotonic dose response  $\mu_1 \leq \mu_1 \leq \dots \leq \mu_k$ , the set of all null hypotheses form a closure since it contains the intersections of all possible subsets of the family of the null hypotheses. In fact, we have  $H_{0k} \supseteq H_{0k-1} \supseteq \dots \supseteq H_{01}$ . Write  $Z_{1-\alpha}$  as the  $1 - \alpha$  percentile point of the standard normal distribution. The following stepdown one-sided testing procedure is a closed procedure and strongly controls the Type I family-wise error probability at level of  $\alpha$ :

- If  $\bar{y}_k - \bar{y}_0 \leq (\frac{1}{n_k} + \frac{1}{n_0})^{\frac{1}{2}}\sigma Z_{1-\alpha}$ , then accept the null hypothesis  $H_{0k}$  and stop the testing procedure;  
else then reject  $H_{0k}$  and go to the next step.
- If  $\bar{y}_{k-1} - \bar{y}_0 \leq (\frac{1}{n_{k-1}} + \frac{1}{n_0})^{\frac{1}{2}}\sigma Z_{1-\alpha}$ , then accept the null hypothesis  $H_{0k-1}$  and reject  $H_{0k}$ , and stop the testing procedure;  
else then reject  $H_{0k}$  and  $H_{0k-1}$  and go to the next step.
- .....
- If  $\bar{y}_1 - \bar{y}_0 \leq (\frac{1}{n_1} + \frac{1}{n_0})^{\frac{1}{2}}\sigma Z_{1-\alpha}$ , then accept  $H_{01}$  and reject  $H_{0k}, H_{0k-1}, \dots, H_{02}$ , and stop the testing procedure;  
else then reject  $H_{0k}, \dots, H_{01}$  and stop the testing procedure.

**Proposition 3.3.1.** *The above stepdown testing procedure strongly controls the family-wise error rate.*

*Proof.*

Suppose that for the  $k$  dose levels, we have the set of true null hypotheses  $T$  such that,  $0 \geq \theta_j \geq \theta_{j-1} \geq \dots \geq \theta_1$ , for  $j \in T$ . Let  $J = \max\{j : j \in T\}$ . Write  $\omega$  as the event that at least one true null hypothesis is rejected, and  $\omega_J$  as the event that  $H_{0J}$  is rejected. By the definition of the stepdown procedure, we know that we have to reject  $H_{0J}$  in order to reject any  $H_{0i}$ ,  $i \in T$ . Therefore we have

$$\begin{aligned} \omega &\subseteq \omega_J \Rightarrow \\ Pr\{\omega\} &\leq Pr\{\omega_J\} \Rightarrow \\ Pr\{\text{Reject any true null hypothesis}\} &\leq Pr\{\text{Reject } H_{0J}\} \\ &= \alpha \end{aligned}$$

Although the above proof does not use the order restriction assumption about the means of dose responses, the construction of the closed family of null hypotheses  $H_{0i}, i \in 1, 2, \dots, k$ , has implicitly used the assumption.

### 3.4 MODIFIED STEPDOWN TESTING PROCEDURES IN ADAPTIVE SAMPLING

In studies using adaptive designs, the usual likelihood based approaches for hypothesis testing and confidence interval are not valid because the data being analyzed are not i.i.d. random samples from the assumed distributions. Modifications of such approaches are needed to perform valid statistical analysis for such data. For example, Jennison and Turnbull [13] in their book discussed approaches to derive statistical inference for the data collected from adaptive clinical trials.

For an adaptive design where the study only stops when evidence of null hypothesis is observed, the actual type I error probability is slightly smaller than the nominal size of the test when the adaptive sampling is ignored, therefore leading to a conservative testing procedure. For this reason, in practice people rarely adjust the test for the adaptive sampling. It is not hard to see that the power of the test also suffers a slight loss too, which people tend to ignore because the loss is usually insignificant. However, in a stepdown testing procedure when we are interested in the power to detect the lowest effective dose level, the loss in the power is exponentially increased when the number of effective dose levels increases, and should not be ignored. For example, suppose in a single test the power of the accurate test is 90% and the power of the conservative (unadjusted) test is 87%. The loss of power is 3.33%, which is ignorable. However, in a stepdown test for a dose finding study, if the lowest effective dose can be as low as the sixth largest dose, the power to find the lowest effective dose becomes  $90\%^6 = 53\%$  and  $87\%^6 = 43\%$ , correspondingly. The loss of power is increased to 19%. It is thus important to perform the stepdown test procedure with accurate size that takes into consideration the adaptive sampling.

In a typical adaptive design for dose study, part of the subjects planned for the dose group will be enrolled first. If the observed drug effect for a given dose group is smaller than a certain pre-specified cut point, the rest of the subjects planned for this dose group will not be enrolled. They can either be allocated to other potentially effective dose groups, or won't be enrolled to save resources for other studies. Suppose instead at dose level  $i$ , that the observed drug effect at the first stage is larger than the cut point, then the rest of the planned subjects for dose  $i$  will be enrolled. Usually only the dose groups fully enrolled will be analyzed as potentially effective dose groups. Also an active control group or a placebo group are continuously enrolled in parallel with the study drug at each dose group. In an adaptive dose study like this, the data in each dose group are not random samples and the

adaptive sampling needs to be incorporated into the data analysis.

Let's assume the same distributions of drug responses discussed in section 3.2. Let the cut point for adaptive sampling be  $\lambda_i$ , i.e., the planned subjects of dose level  $i$  will be fully enrolled if  $\bar{y}_{i1} - \bar{y}_{01} > \lambda_i$ , where  $\bar{y}_{i1}$  and  $\bar{y}_{01}$  are sample means of drug responses of dose group  $i$  and the control group after the first stage of the study. The planned sample size and the sample size at first stage of the study for dose group  $i$  are  $n_i$  and  $n_{i1}$  correspondingly, and the planned sample size and the sample size enrolled at the first stage of the study in the control group are  $n_0$  and  $n_{01}$  correspondingly. Write the mean drug effect observed at the first stage and the second stage of dose group  $i$  as  $\bar{y}_{i1}$  and  $\bar{y}_{i2}$ . Let  $n_{i2} = n_i - n_{i1}$ , then  $\bar{y}_i = \frac{n_{i1}}{n_i} \bar{y}_{i1} + \frac{n_{i2}}{n_i} \bar{y}_{i2}$ . Also write the mean drug effect observed for the subjects in control group enrolled after the first stage of the study as  $\bar{y}_{02}$  with sample size  $n_{02}$ , then we have  $\bar{y}_0 = \frac{n_{01}}{n_0} \bar{y}_{01} + \frac{n_{02}}{n_0} \bar{y}_{02}$ . The  $1 - \alpha$  percentile point  $c_{1-\alpha}$  for drug effect estimate  $\bar{y}_i - \bar{y}_0$  can be obtained by solving

$$\begin{aligned}
& \Pr(\bar{y}_i - \bar{y}_0 > c \text{ and } \bar{y}_{i1} - \bar{y}_{01} > \lambda_i) \\
&= \Pr(\bar{y}_{i1} - \bar{y}_{01} > \lambda_i) \times \Pr(\bar{y}_i - \bar{y}_0 > c | \bar{y}_{i1} - \bar{y}_{01} > \lambda_i) \\
&= \alpha
\end{aligned} \tag{3.1}$$

Under the null hypothesis  $H_{0i}$ , the above probabilities can be expressed separately as

$$\begin{aligned} & \Pr(\bar{y}_i - \bar{y}_0 > c | \bar{y}_{i1} - \bar{y}_{01} > \lambda_i) \\ &= \int_{\lambda_i}^{+\infty} f_{\bar{y}_{i1} - \bar{y}_{01}}(w) \frac{1}{\int_{\lambda_i}^{+\infty} f_{\bar{y}_{i1} - \bar{y}_{01}}(v) dv} \int_{-\infty}^c f_{(\bar{y}_i - \bar{y}_0) | \bar{y}_{i1} - \bar{y}_{01} = w}(z) dz dw \end{aligned}$$

and

$$\Pr(\bar{y}_{i1} - \bar{y}_{01} > \lambda_i) = 1 - \Phi\left(\frac{\lambda_i}{\sigma_{i1}}\right) \quad (3.2)$$

Here  $\Phi$  is the cumulative distribution function of standard normal distribution. The conditional densities  $f_{(\bar{y}_i - \bar{y}_0) | (\bar{y}_{i1} - \bar{y}_{01})}$  can be easily obtained by noticing that under  $H_{0i}$ ,  $f_{\bar{y}_{i1} - \bar{y}_{01}} \sim N(0, \frac{\sigma^2}{n_{i1}} + \frac{\sigma^2}{n_{01}})$ ,  $f_{\bar{y}_i - \bar{y}_0} \sim N(0, \frac{\sigma^2}{n_i} + \frac{\sigma^2}{n_0})$ , and  $\bar{y}_{i1} - \bar{y}_{01}$  and  $\bar{y}_i - \bar{y}_0$  are bivariate normal random variables with covariance  $\frac{\sigma^2}{n_i} + \frac{\sigma^2}{n_0}$ .

### 3.5 A STEPDOWN TESTING PROCEDURE WITH ADAPTIVE SAMPLING INFERENCE

In order to take advantage of the power of stepdown testing procedures and also avoid the bias introduced by adaptive sampling at each dose  $i$ , the following stepdown testing approach is proposed :

- If  $\bar{y}_{k_1} - \bar{y}_0 \leq \left(\frac{1}{n_{k_1}} + \frac{1}{n_0}\right)^{\frac{1}{2}} \sigma c_{1-\alpha}$ , then accept the null hypothesis  $H_{0k_1}$  and stop the testing procedure;
- else then reject  $H_{0k_1}$  and go to the next step.



- If  $\bar{y}_{k_1-1} - \bar{y}_0 \leq \left(\frac{1}{n_{k_1-1}} + \frac{1}{n_0}\right)^{\frac{1}{2}} \sigma c_{1-\alpha}$ , then accept the null hypothesis  $H_{0k_1-1}$  and reject  $H_{0k_1}$ , and stop the testing procedure;  
else then reject  $H_{0k_1}$  and  $H_{0k_1-1}$  and go to the next step.
- . . . . .
- If  $\bar{y}_1 - \bar{y}_0 \leq \left(\frac{1}{n_1} + \frac{1}{n_0}\right)^{\frac{1}{2}} \sigma c_{1-\alpha}$ , then accept the null hypothesis  $H_{01}$  and reject  $H_{0k_1}, \dots, H_{02}$ , and stop the testing procedure;  
else then reject  $H_{0k_1}, H_{0k_1-1}, \dots, H_{01}$  and stop the testing procedure.

Here  $c_{1-\alpha}$  is the critical value derived in section 3.4 by taking into consideration the adaptive design, and  $k_1$  is the number of dose groups fully enrolled.

### 3.6 APPLICATION OF THE MODIFIED STEPDOWN TESTING PROCEDURE IN ADAPTIVE DOSE STUDY WITH A BINARY ENDPOINT

As discussed in the previous chapter, the outcome in clinical studies is often expressed as a binary variable. Usually we can use the approaches of adaptive design and stepdown testing based on normal distribution theory as long as the sample size is moderate or large, and the expected proportion is not very close to 0 or 1. However, in an adaptive study with small to moderate sample size, the sample size at the first stage is usually small, and thus cut points specified based on the continuous normal approximation of testing statistics usually can not be achieved. The cut point is often defined in terms of the number of events instead. In a comparative dose study, the cut points are most conveniently specified as the difference in the number of events between the control group and the dose groups. A cut point specified in this way does not always define a unique cut point when some relative drug efficacy measure

is used. For example, suppose at dose level  $d_i$ , odds ratio  $\frac{p_i \times (1-p_0)}{(1-p_i) \times p_0}$  is defined as the drug effect variable. Use the same notation of sample sizes  $n_i$ ,  $n_{i1}$ ,  $n_0$  and  $n_{01}$  as before, and write the number of event observed in dose  $i$  group and the control group at the first stage as  $m_{i1}$  and  $m_{01}$  correspondingly, and the critical point for  $m_{i1} - m_{01}$  is  $M_i$ . The corresponding critical value  $\lambda_i$  for odds ratio observed at the first stage is

$$\begin{aligned} \lambda_i &= \frac{m_{i1} \times (n_{01} - m_{01})}{(n_{i1} - m_{i1}) \times m_{01}} \\ &= \frac{(m_{01} + M_i) \times (n_{01} - m_{01})}{(n_{i1} - (m_{01} + M_i)) \times m_{01}} \\ &= g(m_{01}, M_i), \end{aligned}$$

Hence  $\lambda_i$  is a function of both  $M_i$  and  $m_{01}$ , and the cut point  $M_i$  does not uniquely define the cut point for the odds ratio. Although the actual corresponding odds ratio cut point can be calculated at the time of testing, the conditional distribution of the observed overall odds ratio can not be derived based on it. The same kind of difficulty will be encountered if the relative risk  $p_i/p_0$  or difference of arcsin square-root of proportions,  $\arcsin p_i^{\frac{1}{2}} - \arcsin p_0^{\frac{1}{2}}$  are used as the drug effect variables. We will propose an approach to deal with this difficulty by using an approximate mixture normal distribution for the odds ratio. The result can be easily applied to relative risk and other similar measurements with minor modifications.

### 3.6.1 Expected conditional distribution approach for stepdown testing in clinical study with binary outcome

Using the above notation, if at dose level  $d_i$  the critical value for adaptive sampling is  $M_i$ , then the corresponding critical value for odds ratio can be expressed as  $\lambda_i = \frac{(m_{01} + M_i) \times (n_{01} - m_{01})}{(n_{i1} - (m_{01} + M_i)) \times m_{01}}$ . Under the null hypothesis, the numbers ( $m_{01}$ ) of event in control group and the number ( $m_{i1}$ ) of event in dose group  $i$  have binomial distributions with common probability  $p$  and

sample sizes  $n_{01}$  and  $n_{i1}$ . Let set  $U_i$  be all the possible values of the odds ratio given sample sizes  $n_{01}$  and  $n_{i1}$  and difference in number of events  $M_i$ . For any value  $\omega \in U_i$ , write the set of  $m_{01}$  as  $V_\omega = \{m : \text{Max}(0, -M_i) \leq m \leq \text{Min}(n_{01}, n_{i1} - M_i) \text{ and } \frac{(m+M_i) \times (n_{01}-m)}{(n_{i1}-(m+M_i)) \times m} = \omega\}$ .

When  $m_{01} \in V_\omega$ , we have

$$m_{i1} - m_{01} \geq M_i \iff \frac{(m_{01} + M_i) \times (n_{01} - m_{01})}{(n_{i1} - (m_{01} + M_i)) \times m_{01}} \geq \omega$$

Therefore the probability of the discrete distribution of the critical value in terms of odds ratio can be evaluated as

$$\begin{aligned} \Pr(\lambda_i = \omega) &= \Pr\{m_{01} \in V_\omega\} \\ &= \frac{1}{G_p} \sum_{m \in V_\omega} \binom{n_{01}}{m} p^m (1-p)^{n_{01}-m}, \end{aligned} \quad (3.3)$$

where

$$G_p = \sum_{m=\text{Max}(0, -M_i)}^{\text{Min}(n_{01}, n_{i1}-M_i)} \binom{n_{01}}{m} p^m (1-p)^{n_{01}-m} \quad (3.4)$$

Since  $p$  is usually unknown, the estimate  $\hat{p} = (m_0 + m_i)/(n_0 + n_i)$  can be plugged into the equation, where  $m_0, n_0, m_i$  and  $n_i$  represent the total number of events and subjects in control group and dose group  $i$ . Thus, approximately,

$$\Pr(\lambda_i = \omega) = \frac{1}{G_{\hat{p}}} \sum_{m \in V_\omega} \binom{n_{01}}{m} \hat{p}^m (1 - \hat{p})^{n_{01}-m} \quad (3.5)$$

Under the null hypothesis, the log odds ratio  $z_{i1} = \log(\text{odd}_{i1})$  has the approximate normal distribution  $N(0, \tilde{\sigma}^2)$ , where  $\tilde{\sigma}^2 = \frac{1}{m_{01}} + \frac{1}{n_{01}-m_{01}} + \frac{1}{m_{i1}} + \frac{1}{n_{i1}-m_{i1}}$ . And  $z_i = \log(\text{odd}_i)$  has the approximate normal distribution  $N(0, \sigma^2)$ , where  $\sigma^2 = \frac{1}{m_0} + \frac{1}{n_0-m_0} + \frac{1}{m_i} + \frac{1}{n_i-m_i}$ . The marginal distribution of the test statistics

$$z_i = \log(\text{odd}_i) = \log \left( \frac{m_i \times (n_0 - m_i)}{(n_i - m_i) \times m_0} \right)$$

can be obtained by the following expected conditional distribution:

$$\begin{aligned} & f(z_i | m_{i1} - m_{0i} \geq M_i) \\ &= \sum_{\omega \in U_i} \Pr \{ \lambda_i = \omega \} f(z_i | z_{i1} \geq \log(\lambda_i)) \end{aligned}$$

And the critical value  $c_{1-\alpha}$  for level  $\alpha$  hypothesis testing of drug effect at dose group  $i$  can be obtained by solving the following:

$$\begin{aligned} & \Pr(m_{i1} - m_{01} \geq M_i \text{ and } z_i \geq C) \\ &= \Pr(m_{i1} - m_{01} \geq M_i) \times \Pr(z_i \geq C | m_{i1} - m_{01} \geq M_i) \\ &= \alpha \end{aligned}$$

Under the null hypothesis  $H_i$ , the above probabilities can be expressed separately as:

$$\begin{aligned} & \Pr(z_i \geq C | m_{i1} - m_{01} \geq M_i) \\ &= \sum_{\omega \in U_i} \Pr(\lambda_i = \omega) \int_{\log(\omega)}^{+\infty} f_{z_{i1}}(w) \frac{1}{\int_{\log(\omega)}^{+\infty} f_{z_{i1}}(v) dv} \int_{-\infty}^{c_{1-\alpha}} f_{z_i | z_{i1}=w}(z) dz dw \end{aligned}$$

and

$$\begin{aligned} & \Pr(m_{i1} - m_{01} \geq M_i) \\ &= \sum_{m_{01}=0}^{\text{Min}(n_{01}, n_{i1}-M_i)} \sum_{m_{i1}=\text{Max}(0, m_{01}+M_i)}^{n_{i1}} \left\{ \binom{n_{01}}{m_{01}} p^{m_{01}} (1-p)^{n_{01}-m_{01}} \right. \\ & \quad \left. \times \binom{n_{i1}}{m_{i1}} p^{m_{i1}} (1-p)^{n_{i1}-(m_{i1})} \right\} \end{aligned}$$

The probability density function  $f_{z_i|z_{i1}}(z)$  of the conditional distribution of  $z_i$  can be easily obtained since the covariance between  $z_i$  and  $z_{i1}$  is known to be approximately equal to  $\sigma^2$ .

### 3.6.2 A Monte Carlo Simulation Approach to Obtain the Critical Values

Solving for the critical values of testing statistics  $z_i$  involves a double integral of the conditional distribution:

$$\int_{\log(\omega)}^{+\infty} f_{z_{i1}}(w) \frac{1}{\int_{\log(\omega)}^{+\infty} f_{z_{i1}}(v) dv} \int_{-\infty}^{c_{1-\alpha}} f_{z_i|z_{i1}=w}(z) dz dw$$

A closed mathematical form of the density function is unavailable. However, we can use Monte Carlo simulation approach to generate a sufficiently large sample from the marginal distribution of  $z_i$ , and estimate the desired percentage point from the empirical distribution. We propose to use a Monte Carlo Composition [30] with Rejective/Acceptance method for our problem. The simulation can be performed in the following three steps to sample from the marginal distribution  $f_{z_i}(z)$ :

1. Draw  $\omega^*$  from  $p_m(\omega)$
2. Draw  $y^*$  from  $f_{z_{i1}}(y)$
3. If  $y^* \geq \omega^*$ , then keep  $y^*$  and go to step 4. Else then go back to step 2
4. Draw  $z^*$  from  $f_{z_i|y^*}(z)$

The point  $(\omega^*, y^*, z^*)$  is an observation from the joint distribution, while  $z^*$  is an observation from the marginal  $f_{z_i}$ . The SAS macro program I wrote to perform the above simulations only needs about 3 to 5 minutes in an artificial trial to compute all the critical values for each of the possible adaptive designs.

**3.6.2.1 Example** The following example of a clinical dose study is used to illustrate the steptest procedure that incorporates adaptive design. Suppose in an adaptive clinical dose trial the treatment outcome is a binomial random variable, which indicates the number of treatment successes. An experimental drug is studied at multiple dose levels for its effect on increasing the success rate. An active control treatment is used, and the drug effect at each dose group is measured as odds ratio relative to the control group. From past data, patients who received the control treatment have been observed to have a 10% success rate following the treatment. The dose study is designed such that 150 patients are enrolled in each of the dose groups and the control group at the first stage. If after the first stage, the number of patients having the adverse outcome at any dose group is no more than that in the control group minus a certain pre-specified number  $\delta$  ( $\delta$  is an integer but could be negative or zero), then at the second stage 350 more subjects will be enrolled into each dose group and the control group.

After completion of the second stage, data from the dose groups that have been fully enrolled will be analyzed for potential drug efficacy. If the dose response can be reasonably assumed to be isotonic, i.e., the drug at higher dose levels is at least as effective as the lower doses, finding the lowest effective dose level is usually desired and a stepdown testing procedure is the most widely used approach. The natural logarithm of the odds ratio is used as the testing statistic and is assumed to be normally distributed. However, the critical value for test statistics at each step has to be adjusted for the adaptive sampling. Table 3.1-3.3 list the one-sided critical values to be used for the test. These critical values are calculated using the Monte Carlo Composition [30] with Rejective/Acceptance method. We assume that the treatment success rate is 10% in the control group, and the treatment success rate is 15% when calculating the power of the test.

The simulation results indicate that the actual Type I error probability for the log odds

Table 8: Adjusted Critical Values for One-sided Test with  $\alpha=0.1$

$\delta$ for Adaptive Design	Probability of 2nd Stage Under $H_0$	Critical Value for Log of Odds Ratio	Simulated Type I Error Probability	Adjusted Power	Unadjusted Power
-8	0.94936	0.26718	0.10351	0.84147	0.83745
-7	0.92599	0.26685	0.10268	0.84186	0.83777
-6	0.89510	0.26509	0.10437	0.84345	0.83651
-5	0.85578	0.26577	0.10221	0.84174	0.83560
-4	0.80752	0.26404	0.10370	0.83995	0.83154
-3	0.75044	0.26351	0.10252	0.83682	0.82866
-2	0.68540	0.26216	0.10234	0.83127	0.82131
-1	0.61399	0.26049	0.09965	0.82240	0.81188
0	0.53847	0.25553	0.09865	0.81404	0.79839
1	0.46153	0.24789	0.09898	0.80172	0.78074
2	0.38601	0.23526	0.09909	0.78525	0.75592
3	0.31460	0.21705	0.09872	0.76048	0.72453
4	0.24956	0.18938	0.10138	0.72816	0.68491
5	0.19248	0.14802	0.10176	0.68632	0.63985
$P_{null} = 0.1 \quad p_a = 0.15 \quad n_1 = 150 \quad n_2 = 350$					

Table 9: Adjusted Critical Values for One-sided Test with  $\alpha=0.05$

$\delta$ for Adaptive Design	Probability of 2nd Stage Under $H_0$	Critical Value for Log of Odds Ratio	Simulated Type I Error Probability	Adjusted Power	Unadjusted Power
-8	0.94936	0.34493	0.052680	0.72500	0.72285
-7	0.92599	0.34351	0.052610	0.72834	0.72338
-6	0.89510	0.34167	0.054145	0.73024	0.72230
-5	0.85578	0.34284	0.052045	0.72729	0.72230
-4	0.80752	0.34119	0.053775	0.72722	0.71933
-3	0.75044	0.34164	0.053080	0.72560	0.71749
-2	0.68540	0.34080	0.053180	0.72064	0.71261
-1	0.61399	0.34108	0.050935	0.71483	0.70699
0	0.53847	0.33896	0.050020	0.70802	0.69719
1	0.46153	0.33400	0.050495	0.70221	0.68482
2	0.38601	0.32659	0.050110	0.69228	0.66633
3	0.31460	0.31527	0.050510	0.68009	0.64285
4	0.24956	0.29924	0.051105	0.66015	0.61170
5	0.19248	0.27558	0.050385	0.63656	0.57675
$P_{null} = 0.1 \quad p_a = 0.15 \quad n_1 = 150 \quad n_2 = 350$					



Table 10: Adjusted Critical Values for One-sided Test with  $\alpha=0.025$

$\delta$ for Adaptive Design	Probability of 2nd Stage Under $H_0$	Critical Value for Log of Odds Ratio	Simulated Type I Error Probability	Adjusted Power	Unadjusted Power
-8	0.94936	0.41194	0.026690	0.59685	0.59685
-7	0.92599	0.41048	0.027710	0.59985	0.59670
-6	0.89510	0.40903	0.027700	0.60472	0.59539
-5	0.85578	0.41005	0.027400	0.60212	0.59607
-4	0.80752	0.40953	0.027420	0.60000	0.59411
-3	0.75044	0.41116	0.027300	0.59722	0.59412
-2	0.68540	0.41078	0.026935	0.59353	0.59037
-1	0.61399	0.41111	0.025995	0.58987	0.58691
0	0.53847	0.41057	0.025200	0.58307	0.58020
1	0.46153	0.40717	0.025480	0.58236	0.57161
2	0.38601	0.40201	0.025650	0.57799	0.55990
3	0.31460	0.39422	0.025560	0.57610	0.54350
4	0.24956	0.38172	0.026200	0.56595	0.52147
5	0.19248	0.36618	0.026420	0.55490	0.49586
$P_{null} = 0.1 \quad p_a = 0.15 \quad n_1 = 150 \quad n_2 = 350$					

ratio using normal approximation is very close to the nominal level in a clinical study with sample size of a typical Phase II trial. The corresponding power under the alternative hypothesis and the probability of subjects at second stage being saved under the null hypothesis can be easily calculated, and these numbers can be used for choosing the best design when the appropriate criteria are given.

## 4.0 OPTIMAL STAGE-WISE ADAPTIVE CLINICAL DOSE STUDY DESIGN

### 4.1 INTRODUCTION

Sequential clinical study designs consist of one or more interim analyses/comparisons in addition to the final analysis/comparison at the end of the study. The study can be stopped if the treatment is found to be inferior or superior to the control at any interim analysis. The use of sequential design is motivated by both the ethical consideration to assign more patients to the effective treatment as soon as possible, and by the economical need to reach the reliable conclusion with the minimum resources. Extensive discussions on various approaches to design sequential studies as well as the theoretical background are available in the books by Jennison and Turnbull [13](2000) and Whitehead [33](1997).

Sequential designs have been widely studied and used in Phase II clinical study involving treatment selection or treatment screening. In drug development process, usually the toxicity and/or tolerability is studied in Phase I study. If the toxicity level is acceptable, potential efficacy is studied in Phase II trials to determine whether a confirmatory Phase III trial, which is often large-scale and very costly, should be conducted. Phase II trial that involves dose response study has become increasingly common, and the parallel multiple dose study is one of the typical Phase II dose designs.

The results of a Phase II drug study often show that a drug candidate believed to be promising lacks sufficient efficacy to warrant further Phase III confirmatory trial. "Futility analysis", which refers to the interim analysis that assesses the probability of the drug being found promising, enables us to stop the study early and allocate the saved resources to more promising studies. In a multi-dose Phase II trial, futility analysis can be conducted at each dose level so that we can allocate more resources to more promising dose groups.

## 4.2 LITERATURE REVIEW ON ADAPTIVE PHASE II CLINICAL STUDY DESIGNS

Simon [23](1989) proposed an approach to design the optimal two-stage single-arm phase II study. It allows the study to stop after the first stage if there is no sufficient evidence of drug activity. Minimum expected sample size and/or minimum maximum sample size are used to choose the optimal design. Chen [7] (1997) proposed the extension of the design to three-stage. Hanfelt [11](1999) extended Simon's design to phase II studies where choice of dosage is not clear. Hanfelt's design has two stages at each dose, and permits one dose escalation if the drug effect can't be verified after first and/or second stage of the initial dose. Different optimization criteria are discussed, including expected sample size, median or mode sample size, and minimum maximum sample size.

Feder et al [10](1991) discussed a two-stage adaptive dose allocation optimal design for parametric dose response estimation. At the first stage, subjects are allocated to dose levels such that the parameters can be best estimated based on the prior assumption. At the second stage, the data from the first stage are used to update the function form and the subjects allocation is modified accordingly.

Thall [31][32](1988 and 1989) studied a two-stage design with binary outcome that selects the best treatment at the first stage, and conduct the second stage study only on the selected treatment if the observed treatment effect crosses some pre-specified boundary in first stage. He proposed to use two quantities, marginal improvement  $\delta_1$ , and significant improvement  $\delta_2$  to derive the Least Favorable Configuration (LFC), and to calculate the power and Type I error probability based on LFC. The optimal design is obtained to minimize the weighted average of the sample sizes of the two stages of study. Schaid, Wieand and Therneau [21](1990) proposed a similar design when the endpoint variable is a survival time. Stallard [26](2003) extended the work by Thall and Schaid by using the efficiency score as the test statistic, and incorporating sequential design in studying the single experimental treatment at the second stage after the best treatment has been selected at the first stage.

Leung [16](2001) incorporated the usually assumed monotone relationship between toxicity and dosage into the design. He suggested a sequential dose finding study that performs isotonic regression after each stage of the study, and uses the isotonic regression estimate of the drug effect as a test statistic to decide whether to stop the study or to continue to the next dose level.

Pallay [20](2001) discussed a Phase II two-dose futility study using a decision analytic approach. Two-stage sequential design is used at each dose. Cost and gain functions are pre-specified and the study design is optimized based on the eventual profit of the drug development. Stallard [27] [26] (1998 and 2003) proposed to use decision-theoretic design based on more general gain function of drug effect and subjects allocation. The opportunity cost of not spending resources on other competing treatments is also considered.

Previous studies of adaptive design for Phase II clinical trial have been limited to one or more of the following areas, therefore can not be readily applied to the kind of Phase II dose

study we have described: 1. Frequently, usually in early stage toxicity study, parametric dose response estimation is desired. In those cases, the optimal design is aimed to achieve the minimum variance in parameter estimation(s). 2. In a Phase II study when multiple treatments are compared to a control group to select potentially effective treatment(s) to enter a Phase III study, the monotone assumption of effect exists in dose study either does not exist or is not taken in consideration into the study design. 3. Usually only a continuous response with a normal distribution is considered. In case of binary or categorical responses, certain types of normal approximation and/or data transformation are used without validity and sensitivity analysis. 4. Most study designs simply use the minimum expected sample size or min-max sample size as the optimization criteria. However, sometimes the total resources to be expended has already been decided and efficiency of resource allocation is the study design objective. 5. Extensive research has been concentrated on Phase II cancer study, where the tumor reduction probability is examined to assess the potential therapeutic effect of new chemical compounds. Therefore, the studies are singled-armed and are not comparative in nature.

In practical clinical studies, the study process and design are in general constrained by scientific and clinical considerations and operation logistics. A general statistically optimal design doesn't always find its way into the application. In this chapter we will propose a study design that satisfies the constraints of certain type of dose escalation studies and also achieves the maximum statistical efficiency under the study constraints.

### 4.3 AN ADAPTIVE STUDY DESIGN FOR DOSE ESCALATION STUDY

The study design we will discuss here is aimed to achieve optimal efficiency for a comparative dose escalation Phase II clinical study similar to the cardiovascular clinical study mentioned in the first chapter. Such a dose escalation study is commonly used in clinical studies involving severe medical conditions. The efficacy of the drug is often measured using a binary variable that indicates the treatment success of the target disease or the occurrence of the target adverse outcome. Usually two to six pre-specified dose levels of the experimental drug are studied and isotonic dose response can be assumed. An active control arm that uses the standard treatment is used, and the drug effect is measured by a relative measurement such as difference of proportions, relative risk or odds ratio.

#### 4.3.1 Proposed Study Design

Suppose there are  $k$  dose levels,  $d_1, d_2, \dots, d_k$ , and an active control arm  $d_0$ , and the measure of drug effect at the control group and each dose level is normally distributed with the same known variance  $\sigma^2$ , and with means  $\mu_0, \mu_1, \dots, \mu_k$ , respectively. The study starts from the lowest dose level, and escalates to the second lowest dose level after the study enrollment in the lowest doses is finished and study data has been collected and analyzed. The study continues in this way until the highest dose level is studied, or the study is terminated due to safety concerns. To keep the treatment allocation blinded and also the control population comparable, the subjects in the control arm are continuously enrolled along with subjects to be allocated to each dose level. Randomization by block is used to maintain the desired ratio of the numbers of subjects between control and treatment.

I propose to use a two-stage adaptive enrollment design at each dose step. The enrollment design is aimed to use the subjects more efficiently by performing a futility assessment after

the first stage enrollment, and skipping the second stage enrollment if the observed drug effect of the dose group is smaller than a pre-specified cut point. The saved second stage subjects of the abandoned dose group are then equally allocated to the remaining higher-dose groups. The number of subjects at first stage enrollment is the same at every dose step. Suppose at a dose step, say dose  $i$ , the observed drug effect after the first stage enrollment is larger than the cut point  $\lambda_i$ , then the second stage enrollment will proceed. After the dose step  $i$  is completed, all the higher dose level will be fully enrolled without first stage futility assessment.

Suppose the total planned number of subjects for  $k$  dose groups and the control group is  $N$ . At the first stage enrollment of each dose step only  $n_1$  ( $n_1 < \frac{N}{k+1}$ ) subjects are enrolled. I propose the following subject allocation scheme which ensures that the control group and all the dose groups that pass the first stage screen have the same number of subjects. Also suppose that after the first stage of the adaptive enrollment, the drug effect for the control group and dose group  $i$ ,  $i = 1, 2, \dots, k$ , is estimated using the observed sample means  $\bar{x}_{0i}$  and  $\bar{x}_{i1}$  respectively. Write the cut point at dose group  $i$  as  $\lambda_i$ . The dose escalation study starts from the lowest dose group:

1. Enroll  $n_1$  subjects in the lowest dose group and  $\frac{N}{2k(k+1)}$  subjects in the control group.

Calculate the drug effect estimate in the control group and the lowest dose group,  $\bar{x}_{01}$  and  $\bar{x}_{11}$ .

- If  $\bar{x}_{11} - \bar{x}_{01} \leq \lambda_1$  then stop enrollment for the lowest dose group and go to the second lowest dose group (Step 2).
- Else if  $\bar{x}_{11} - \bar{x}_{01} > \lambda_1$ , then enroll additional  $\frac{N}{k+1} - n_1$  subjects to the lowest dose group, and additional  $\frac{N}{2k(k+1)}$  subject to the control group. In each of the higher dose steps, enroll  $\frac{N}{k+1}$  and  $\frac{N}{k(k+1)}$  subjects to the dose group and the control group, respectively. (Study enrollment finished)



2. Enroll  $n_1$  subjects in the second lowest dose group, and  $\frac{N-n_1}{2k(k-1)} - \frac{N}{2k(k+1)}$  subjects in the control group. Calculate the drug effect estimate in the control group (use all data) and the second lowest dose group,  $\bar{x}_{02}$  and  $\bar{x}_{21}$ .

- If  $\bar{x}_{21} - \bar{x}_{02} \leq \lambda_2$  then stop enrollment for the second lowest dose group and go to the third lowest dose group (Step 3).
- Else if  $\bar{x}_{21} - \bar{x}_{02} > \lambda_2$ , then enroll additional  $\frac{N-n_1}{k} - n_1$  subjects to the second lowest dose group, and additional  $\frac{N-n_1}{2k(k-1)}$  subject to the control group. In each of the higher dose steps, enroll  $\frac{N-n_1}{k}$  and  $\frac{N-n_1}{k(k-1)}$  subjects to the dose group and the control group, respectively. (Study enrollment finished)

.....

I. Enroll  $n_1$  subjects in dose group  $i$ , and  $\frac{N-(i-1)n_1}{2(k-i+1)(k-i+2)} - \frac{N-(i-2)n_1}{2(k-i+2)(k-i+3)}$  subject in the control group. Calculate the drug effect estimate in the control group (use all the data collected in the control group) and the second highest dose group,  $\bar{x}_{0i}$  and  $\bar{x}_{i1}$ .

- If  $\bar{x}_{i1} - \bar{x}_{0i} \leq \lambda_i$  then stop enrollment for dose group  $i$  and go to dose group  $i + 1$ .
- Else if  $\bar{x}_{i1} - \bar{x}_{0i} > \lambda_i$ , then enroll additional  $\frac{N-(i-1)n_1}{k-i+1} - n_1$  subjects to dose group  $i$ , and additional  $\frac{N-(i-1)n_1}{2(k-i+1)(k-i+2)}$  subject to the control group. In the highest dose step, enroll  $\frac{N-(i-1)n_1}{k-i+1}$  and  $\frac{N-(i-1)n_1}{(k-i+1)(k-i+2)}$  subjects to the dose group and the control group, respectively. (Study enrollment finished)

.....

K. In the highest dose step, enroll  $\frac{N-(k-1)n_1}{2}$  and  $\frac{N-(k-1)n_1}{2} - \frac{N-(k-2)n_1}{12}$  subjects to the dose group and the control group, respectively. (Study enrollment finished)

Dose groups that stop enrollment after the first stage are considered not having sufficient evidence of efficacy to be further studied. Suppose at dose level  $d_j$ , we observed that  $\bar{x}_{j1} - \bar{x}_{0j} > \lambda_j$  for the first time, and therefore enroll the remaining subjects such that the control

group and the dose groups at or above dose level  $j$  have the same sample size. The final estimates of drug effect for all the dose groups at or above dose level  $d_j$  are correlated with  $\bar{x}_{j1} - \bar{x}_{0j}$  because they use the common control group. Since the distribution of  $\bar{x}_{j1} - \bar{x}_{0j}$  is truncated by the adaptive sampling procedure, the marginal distributions of the drug effect estimates for all the higher dose group are all biased. Therefore the statistical inferences we use to analyze the study data need to be adjusted.

I will describe the stepdown testing procedure that incorporates the adaptive enrollment in a three-dose dose escalation study. Write the drug effect estimate for dose group  $i$  as  $\bar{x}_i - \bar{x}_0$ ,  $i = 1, 2$  or  $3$ . Also write the event that the null hypothesis  $H_{i0} : \mu_i - \mu_0 \leq 0$  is rejected when  $H_{i0}$  is in fact true as  $S_i$ , i.e.,  $S_i$  is the event the a Type I error occurs in the testing procedure at dose level  $i$ . When  $i = 1$ :

$$\begin{aligned}
& Prob(S_1) \\
&= Prob(\bar{x}_1 - \bar{x}_0 > C_1 \text{ and } \bar{x}_{11} - \bar{x}_{01} > \lambda_1 | \mu_1 \leq \mu_0) \\
&\leq Prob(\bar{x}_1 - \bar{x}_0 > C_1 \text{ and } \bar{x}_{11} - \bar{x}_{01} > \lambda_1 | \mu_1 = \mu_0) \\
&= Prob(\bar{x}_1 - \bar{x}_0 > C_1 | \bar{x}_{11} - \bar{x}_{01} > \lambda_1, \mu_1 = \mu_0) \times Prob(\bar{x}_{11} - \bar{x}_{01} > \lambda_1 | \mu_1 = \mu_0)
\end{aligned}$$

Therefore the derivation of the critical value  $C_1$  is reduced to the problem we have discussed in chapter 3 (3.1) by setting the last probability to  $\alpha$  and solving for  $C_1$ . When  $i = 2$ , the event  $S_2$  can be expressed as:

$$\begin{aligned}
S_2 = & \{ \bar{x}_{11} - \bar{x}_{01} \leq \lambda_1, \bar{x}_{21} - \bar{x}_{02} \leq \lambda_2, \bar{x}_2 - \bar{x}_0 > C_{21} \mid \mu_2 \leq \mu_0 \} \\
& \cup \{ \bar{x}_{11} - \bar{x}_{01} > \lambda_1, \bar{x}_2 - \bar{x}_0 > C_{22} \mid \mu_2 \leq \mu_0 \}
\end{aligned}$$

Since the two subsets of  $S_2$  are mutually exclusive, we have

$$\begin{aligned}
& Prob(S_2) = Prob(\bar{x}_{11} - \bar{x}_{01} \leq \lambda_1, \bar{x}_{21} - \bar{x}_{02} \leq \lambda_2, \bar{x}_2 - \bar{x}_0 > C_{21} \mid \mu_2 \leq \mu_0) \\
& \quad + Prob(\bar{x}_{11} - \bar{x}_{01} > \lambda_1, \bar{x}_2 - \bar{x}_0 > C_{22} \mid \mu_2 \leq \mu_0) \\
& \leq Prob(\bar{x}_{21} - \bar{x}_{02} > \lambda_2, \bar{x}_2 - \bar{x}_0 > C_{21} \mid \mu_2 = \mu_0) \\
& \quad + Prob(\bar{x}_{11} - \bar{x}_{01} > \lambda_1, \bar{x}_2 - \bar{x}_0 > C_{22} \mid \mu_2 = \mu_1 = \mu_0) \\
& = Prob(\bar{x}_2 - \bar{x}_0 > C_{21} \mid \bar{x}_{21} - \bar{x}_{02} > \lambda_2, \mu_2 = \mu_0) \times Prob(\bar{x}_{21} - \bar{x}_{02} > \lambda_2 \mid \mu_2 = \mu_0) \\
& \quad + Prob(\bar{x}_2 - \bar{x}_0 > C_{22} \mid \bar{x}_{11} - \bar{x}_{01} > \lambda_1, \mu_1 = \mu_2 = \mu_0) \\
& \quad \times Prob(\bar{x}_{11} - \bar{x}_{01} > \lambda_1 \mid \mu_1 = \mu_0)
\end{aligned}$$

I propose to use the Bonferroni method to assign Type error rate  $\frac{\alpha}{2}$  to each of the two probabilities at the last step of above inequality. Again the critical values  $C_{21}$  and  $C_{22}$  can be solved by using the simulation approach I proposed in Chapter 3. Under the proposed study design, the testing at the highest dose level (dose 3) is always performed irrespective of the results of futility assessments at early dose steps. Although the results of the futility assessments affect the sample size of the dose effect estimates at dose level 3, the correlation between  $\bar{x}_3 - \bar{x}_0$  and the first stage estimates at earlier dose levels,  $\bar{x}_{11} - \bar{x}_{01}$  and  $\bar{x}_{21} - \bar{x}_{02}$ , is usually small. This is easy to understand since the only common data used is the first stage data in the control group, which is usually a fraction of the sample size at the last dose group. Therefore I propose to use  $\frac{1}{\sqrt{2/n}}\sigma Z_\alpha$  as the critical value of the testing at dose level 3, which is essential the usual Z-test. Here  $Z_\alpha$  is the  $\alpha$  percentile of standard normal distribution,  $\sigma$  is the common standard deviation of drug response, and  $n$  is the final sample size in the control group and each dose group that is fully enrolled. The simulation study I performed shows that the actual Type I error is very close the nominal level  $\alpha$ .

When inference on simultaneous confidence intervals (SCI) is desired, the modified max-min procedure can be used to derive the simultaneous conference intervals of the correlated

dose effects, and a similar adjustment for the adaptive enrollment at dose  $i$  can be made. The modification for adaptive enrollment using SCI will be discussed later in this chapter.

The modified stepdown testing procedure or the one-sided modified simultaneous confidence intervals approach to find the effective dose levels are the two approaches most widely used to analyze comparative multi-dose study. I propose a stage-wise optimization method to obtain the cut points for adaptive sampling at each dose level when these two analysis approaches are used. However, the optimization method can also be applied when other analysis approaches are used or these two analysis approaches are modified.

#### **4.3.2 Optimal Adaptive Design When Stepdown Testing Procedure is Used to Analyze the Study Data**

The goal of using the cut points to assess the drug futility in this adaptive design is to make the best decision on whether the second stage enrollment at a dose step should be proceeded based on the observed drug effect from the first stage data. In order to discuss the optimization approach to decide these cut points, I'll first define the design criteria that I will use to evaluate different cut points.

In a Phase II dose study, one desired property of a study design is to detect the effectiveness of the experimental drug with as much power as possible, which is also the objective of any Phase II clinical study. In a dose study, if the efficacy exists at multiple dose levels, we also wish to identify the lowest effective dose level that is studied. These two competing objectives require somewhat different subject allocation strategies. To have good power to detect the effectiveness of a experimental drug, we should assign as many subjects to the highest dose level as possible, since the highest dose level is the most effective under isotonic dose response assumption. On the other hand, to identify the lowest effective dose level, we

need to assign more subjects to all of the effective dose groups. Average power to detect all the effective dose groups is often used as one of the criteria to evaluate the performance of dose study analysis approaches. Tamhane [29] used this criteria along with other quantities to compare the efficiency of various step testing procedures to analyze dose study data. In order to choose the study design that achieves the optimal balance between the two objectives of a dose study, we propose to use a weighted average power to find all the effective dose groups as the study design criteria. Suppose we have  $k$  dose levels  $d_1, d_2, \dots, d_k$ ,  $r$  effective dose levels  $d_{k-r+1}, d_{k-r+2}, \dots, d_k$ . Write the event that dose  $i$  ( $k-r+1 \leq i \leq k$ ) and all higher dose levels being identified to be effective as  $I_i$ , for a specific study design, the design criteria  $G_r$  is defined as

$$G_r = \sum_{i=k-r+1}^k \gamma_i Pr(I_i|s) / \sum_{i=k-r+1}^k \gamma_i, \quad \gamma_{k-r+1} \geq \gamma_{k-r+2} \geq \dots \geq \gamma_k$$

The relative values of  $\gamma_i$ 's should be decided based on the nature of the therapeutic area studied by the dose study and the corresponding study goal. Specifically,  $\gamma_1 = \gamma_2 = \dots = \gamma_k$  is used in cases where the target disease is extremely severe and there is no alternative effective treatment. In such cases, the goal is to identify the tolerable effective dose level as soon as possible, and identifying lower effective dose levels provides little additional benefit as long as some higher effective dose level(s) can be identified. On the other hand, weights  $\gamma_1, \gamma_2, \dots, \gamma_k$  that reflect steep decrease in  $G_s$  when the number of identified effective dose levels decreases can be used in cases where the degree of effectiveness is less important, and higher doses often involve serious side effects, difficulty in administration and/or much higher cost. In these cases, finding the lowest effective dose level is the most important objective. Appropriate selection of  $\gamma_i$ 's can achieve the desired balance when both objectives are important.

I assume that the drug response variables in the control group and each dose group

are normally distributed, with means of the distributions as  $\mu_0, \mu_1, \dots, \mu_k$  correspondingly. The dose level  $i$  is defined to be effective if  $\mu_i > \mu_0$ , and ineffective otherwise. Let  $\delta$  be the smallest meaningful drug effect that is measured as the difference between drug responses of the control group and each of the dose groups, and we will be interested in the power of the analysis procedure to detect the dose levels that have drug effect higher than  $\delta$  ( $\mu_i > \mu_0 + \delta$ ). The use of smallest meaningful effect  $\delta$  is also statistically useful since the power to distinguish between effective and ineffective dose levels can not be analyzed if  $\mu_i$  and  $\mu_0$  are allowed to be arbitrarily close to each other [31].

**4.3.2.1 A Conservative Stage-Wise Optimization Approach Under Least Favorable Configurations** Optimization of study design parameters are usually achieved under a set of possible drug effect configurations. For example, Simon [23], Hanfelt [11] and Chen [7] provided the optimal cut points for their adaptive designs under a list of possible drug effect configurations in the treatment group and/or the control group. However, in a Phase II study with multiple dose groups, the dose response of the experimental drug is usually not clear, and it is unrealistic to specify a set of possible drug effects at each dose with any confidence. Thall [31] [32] in his proposed adaptive 2-stage design derived the optimal cut points under the least favorable configuration (LFC). The LFC in Thall's study design is the drug effect configuration under which the overall power to find the effective drug group(s) achieves the minimal. This is justified by the fact that in a well studied therapeutic area in treating serious diseases, one usually expects the improvement of treatment effect of any experimental medicine over the standard therapy to be only marginal, if any.

I propose to use a conservative stepwise optimization approach for the adaptive study design here. The least favorable configurations are defined at each dose step such that the probability of making the unfavorable enrollment decision is the highest, and the weighted

average power to find all effective dose groups is the smallest. Suppose the study is at dose step  $i$ , and we have the observed treatment effect after the first stage enrollment  $\bar{x}_{i1} - \bar{x}_{0i}$ . The unfavorable enrollment decision is made if we observe  $\bar{x}_{i1} - \bar{x}_{0i} > \lambda_i$  when in fact  $\mu_i \leq \mu_0$ , or if we observe  $\bar{x}_{i1} - \bar{x}_{0i} \leq \lambda_i$  when in fact  $\mu_i \geq \mu_0 + \delta$ . I define the opportunity loss of using the adaptive design as the decrease in weighted average power when the unfavorable enrollment decision is made. The optimal value of the cut point  $\lambda_i$  is derived such that under least favorable configurations the maximum opportunity loss achieves its minimum.

Define the following two dose response configurations at dose step  $i$ ,  $i = 1, 2, \dots, k - 1$ ,

- $LFC_1^i$ :  $\mu_k = \mu_{k-1} = \dots = \mu_{i+1} = \mu_0 + \delta$ ;  $\mu_i = \mu_0$
- $LFC_2^i$ :  $\mu_i = \mu_{i+1} = \dots = \mu_k = \mu_0 + \delta$

I'll prove as following that under each of the two dose response configurations at dose  $i$ , the probability of making a unfavorable enrollment decision after the interim look is the biggest, while in the same time the expected weighted average power is the smallest.

**Proposition 4.3.1.** *Suppose the drug responses have the normal distributions as described earlier in this Chapter. At dose level  $i$ ,  $i = 1, 2, \dots, k - 1$ , define*

$$p_{i1} = Prob(\bar{x}_{i1} - \bar{x}_{0i} > \lambda_i \mid \mu_i \leq \mu_0)$$

$$p_{i2} = Prob(\bar{x}_{i1} - \bar{x}_{0i} \leq \lambda_i \mid \mu_i \geq \mu_0 + \delta)$$

1.  $p_{i1}$  or  $p_{i2}$  achieves the maximum under  $LFC_1$  or  $LFC_2$ , correspondingly.
2. The weighted average power to identify all effective dose levels ( $G$ ) achieves its minimum under  $LFC_1$  or  $LFC_2$ , correspondingly.

*Proof.*

1.

$$\begin{aligned}
p_{i1} &= Prob(\bar{x}_{i1} - \bar{x}_{01} > \lambda_i \mid \mu_i \leq \mu_0) \\
&= Prob\left(\frac{(\bar{x}_{i1} - \bar{x}_{01}) - (\mu_i - \mu_0)}{\sqrt{\sigma^2/n_{i1} + \sigma^2/n_{01}}} > \frac{\lambda_i - (\mu_i - \mu_0)}{\sqrt{\sigma^2/n_{i1} + \sigma^2/n_{01}}} \mid \mu_i \leq \mu_0\right) \\
&= Prob\left(Z > \frac{\lambda_i - (\mu_i - \mu_0)}{\sqrt{\sigma^2/n_{i1} + \sigma^2/n_{01}}} \mid \mu_i \leq \mu_0\right) \quad (Z \sim N(0, 1)) \\
&\leq Prob\left(Z > \frac{\lambda_i - (\mu_i - \mu_0)}{\sqrt{\sigma^2/n_{i1} + \sigma^2/n_{01}}} \mid \mu_i = \mu_0\right) \\
&= \Phi\left(\frac{\lambda_i}{\sqrt{\sigma^2/n_{i1} + \sigma^2/n_{01}}}\right)
\end{aligned}$$

Similarly, we can also show that

$$\begin{aligned}
p_{i2} &= Prob(\bar{x}_{i1} - \bar{x}_{01} \leq \lambda_i \mid \mu_i \geq \mu_0 + \delta) \\
&= Prob\left(\frac{(\bar{x}_{i1} - \bar{x}_{01}) - (\mu_i - \mu_0)}{\sqrt{\sigma^2/n_{i1} + \sigma^2/n_{01}}} \leq \frac{\lambda_i - (\mu_i - \mu_0)}{\sqrt{\sigma^2/n_{i1} + \sigma^2/n_{01}}} \mid \mu_i \geq \mu_0 + \delta\right) \\
&= Prob\left(Z \leq \frac{\lambda_i - (\mu_i - \mu_0)}{\sqrt{\sigma^2/n_{i1} + \sigma^2/n_{01}}} \mid \mu_i \geq \mu_0 + \delta\right) \\
&\leq Prob\left(Z \leq \frac{\lambda_i - (\mu_i - \mu_0)}{\sqrt{\sigma^2/n_{i1} + \sigma^2/n_{01}}} \mid \mu_i = \mu_0 + \delta\right) \\
&= 1 - \Phi\left(\frac{\lambda_i - \delta}{\sqrt{\sigma^2/n_{i1} + \sigma^2/n_{01}}}\right)
\end{aligned}$$

2. Suppose there are  $r$  effective dose levels. By the assumption of isotonic dose response, we have the effective dose groups as  $k - r + 1, k - r + 2, \dots, k$ . When a stepdown testing procedure is used to identify the effective dose levels, we have

$$\begin{aligned}
G_r &= \sum_{i=k-r+1}^k \gamma_i Pr(I_i) / \sum_{i=k-r+1}^k \gamma_i \\
&= \sum_{i=k-r+1}^k \gamma_i Pr(\text{Reject } H_{k0}, H_{k-10}, \dots, H_{i0} \mid \mu_k \geq \dots \geq \mu_i \geq \mu_0 + \delta) / \sum_{i=k-r+1}^k \gamma_i \\
&\geq \sum_{i=k-r+1}^k \gamma_i Pr(\text{Reject } H_{k0}, H_{k-10}, \dots, H_{i0} \mid \mu_k = \dots = \mu_i = \mu_0 + \delta) / \sum_{i=k-r+1}^k \gamma_i
\end{aligned} \tag{4.1}$$



Since  $Pr(I_i) \leq Pr(I_i) \leq \dots \leq Pr(I_k)$ , it is obvious that for, any given sampling scheme,  $G_r$  achieves the minimum when the number of effective groups  $r$  is as large as possible.

Under any given drug effect configuration, the weighted expected power is a function of the cut point  $\lambda_i$ . At dose level  $i$ , write the opportunity loss under  $LFC_1^i$  and  $LFC_2^i$  as  $L_1^i$  and  $L_2^i$  respectively, they can be expressed as:

$$\begin{aligned} L_1^i &= G_{k-i}(\text{No 2nd Stage Enrollment} \mid LFC_1^i) - G_{k-i}(\lambda_i \mid LFC_1^i) \\ L_2^i &= G_{k-i+1}(\text{Full Enrollment} \mid LFC_2^i) - G_{k-i+1}(\lambda_i \mid LFC_2^i) \\ L^i(\lambda_i) &= \max(L_1^i, L_2^i) \end{aligned}$$

The weighted average power  $G_{k-i}(\text{No 2nd Stage Enrollment} \mid LFC_1^i)$  is equivalent to setting the cut point  $\lambda_i$  at  $+\infty$  when the dose response configuration is under  $LFC_1^i$ , and the weighted average power  $G_{k-i+1}(\text{Full Enrollment} \mid LFC_2^i)$  is equivalent to setting the cut point  $\lambda_i$  at  $-\infty$  when the dose response configuration is under  $LFC_2^i$ . The optimal cut point  $\lambda_i$  is chosen such that the maximum opportunity loss  $L_i$  at dose step  $d_i$  achieves the minimum. As the dose step moves up, the sample size available increases and the number of dose groups higher than the dose step decreases. The optimization procedure is then performed at each dose step to derive the cut points for the adaptive design.

**4.3.2.2 A Back-Induction Algorithm to Solve for the Optimal Cut Points** If after the first stage enrollment of dose step  $i$ ,  $i = 1, 2, \dots, k - 1$ , we observe  $\bar{x}_{i1} - \bar{x}_{0i} > \lambda_i$ , then the second stage enrollment will be proceeded, and all the higher dose group will be enrolled with the same sample size. The weighted average power for finding the effective dose groups under a given dose response configuration can be calculated. However, we

observe  $\bar{x}_{i1} - \bar{x}_{0i} \leq \lambda_i$ , the study escalates to dose step  $i + 1$ , and adaptive enrollment will be implemented again. The calculation of the weighted average power of the study then will involve the cut point  $\lambda_{i+1}$ . Actually, the cut points for all the higher dose steps  $\lambda_{i+1}, \lambda_{i+2}, \dots, \lambda_{k-1}$  need to be known to calculate the weighted average power. Therefore, I propose to use a back-induction algorithm to derive the optimal cut points  $\lambda_1, \lambda_2, \dots, \lambda_{k-1}$  for the adaptive dose escalation study.

I will derive the proposed back-induction optimization algorithm in a dose escalation study with three dose groups. Suppose the total sample size for the dose study is  $N$ , and sample size for the first stage enrollment at each dose step is  $n_1$ . Using the treatment allocation scheme presented previously in the chapter, the sample size of each treatment group under different enrollment decision can be expressed as:

- $\bar{x}_{11} - \bar{x}_{01} < \lambda_1$  and  $\bar{x}_{21} - \bar{x}_{02} < \lambda_2$ :  $N_3 = (N - 2n_1)/2$
- $\bar{x}_{11} - \bar{x}_{01} < \lambda_1$  and  $\bar{x}_{21} - \bar{x}_{02} \geq \lambda_2$ :  $N_2 = (N - n_1)/3$
- $\bar{x}_{11} - \bar{x}_{01} \geq \lambda_1$ :  $N_1 = N/3$

Write the event that lowest effective dose level 1, 2 or 3 is identified using the stepdown testing procedure as  $I_1, I_2$  and  $I_3$ , correspondingly, and the weights of the weighted average power are  $\gamma_1, \gamma_2$  and  $\gamma_3$ . The optimal cut points are derived using the 2-step procedure:

1. Suppose  $\bar{x}_{11} - \bar{x}_{01} < \lambda_1$  is observed, therefore the second stage enrollment of dose step 1 is skipped. Observed drug effect of dose level 2 after the first stage enrollment is  $\bar{x}_{21} - \bar{x}_{02}$ . Using cut point  $\lambda$ , the expected weighted average power under  $LF C_1^2 (\mu_2 = \mu_0)$

and  $\mu_3 = \mu_0 + \delta$ ) and  $LFC_2^2$  ( $\mu_2 = \mu_3 = \mu_0 + \delta$ ) can be calculated as

$$\begin{aligned}
E(G_1^2|\lambda) &= \text{Prob}(\bar{x}_{21} - \bar{x}_{02} \geq \lambda \mid \mu_2 = \mu_0) \times \text{Prob}(I_3|N_2) \\
&\quad + \text{Prob}(\bar{x}_{21} - \bar{x}_{02} < \lambda \mid \mu_2 = \mu_0) \times \text{Prob}(I_3|N_2) \\
E(G_2^2|\lambda) &= \text{Prob}(\bar{x}_{21} - \bar{x}_{02} \geq \lambda \mid \mu_2 = \mu_0 + \delta) \\
&\quad \times (\gamma_2 \text{Prob}(I_2|N_2) + \gamma_3 \text{Prob}(I_3|N_2)) / (\gamma_2 + \gamma_3) \\
&\quad + \text{Prob}(\bar{x}_{21} - \bar{x}_{02} < \lambda \mid \mu_2 = \mu_0 + \delta) \times \text{Prob}(I_3|N_3) \frac{\gamma_3}{\gamma_2 + \gamma_3}
\end{aligned}$$

The weighted average power when "correct" enrollment decision is made, i.e., skip the second stage enrollment at dose step 2 under  $LFC_1^2$ , or proceed to the second stage enrollment under  $LFC_2^2$ , can be calculated as

$$\begin{aligned}
\tilde{G}_1^2 &= \text{Prob}(I_3|N_3) \\
\tilde{G}_2^2 &= (\gamma_2 \text{Prob}(I_2|N_2) + \gamma_3 \text{Prob}(I_3|N_2)) / (\gamma_2 + \gamma_3)
\end{aligned}$$

The optimal cut point at dose step 2  $\lambda_2$  can be derived by minimize the maximum opportunity loss as following:

$$L_2(\lambda) = \text{Max}(\tilde{G}_1^2 - E(G_1^2|\lambda), \tilde{G}_2^2 - E(G_2^2|\lambda))$$

2. At the lowest dose step (dose 1), observed drug effect after the first stage enrollment is  $\bar{x}_{11} - \bar{x}_{01}$ . Notice that the drug effect configuration for dose level 2 and 3 under least favorable configuration  $LFC_1^1$  is the same as  $LFC_2^2$  ( $\mu_2 = \mu_3 = \mu_0 + \delta$ ). Using the cut

point  $\lambda$  at dose step 1, the expected weighted average power under  $LFC_1^1$  ( $\mu_1 = \mu_0$  and  $\mu_3 = \mu_2 = \mu_0 + \delta$ ) and  $LFC_2^1$  ( $\mu_1 = \mu_2 = \mu_3 = \mu_0 + \delta$ ) can be calculated as

$$\begin{aligned}
E(G_1^1|\lambda, \lambda_2) &= \text{Prob}(\bar{x}_{11} - \bar{x}_{01} \geq \lambda | \mu_1 = \mu_0) \times (\gamma_2 \text{Prob}(I_2|N_1) + \gamma_3 \text{Prob}(I_3|N_1)) \\
&\quad / (\gamma_2 + \gamma_3) + \text{Prob}(\bar{x}_{11} - \bar{x}_{01} < \lambda | \mu_1 = \mu_0) \times E(G_2^2|\lambda_2) \\
E(G_2^1|\lambda, \lambda) &= \text{Prob}(\bar{x}_{11} - \bar{x}_{01} \geq \lambda | \mu_1 = \mu_0 + \delta) \\
&\quad \times (\gamma_1 \text{Prob}(I_1|N_1) + \gamma_2 \text{Prob}(I_2|N_1) + \gamma_3 \text{Prob}(I_3|N_1)) / (\gamma_1 + \gamma_2 + \gamma_3) \\
&\quad + \text{Prob}(\bar{x}_{11} - \bar{x}_{01} < \lambda | \mu_1 = \mu_0 + \delta) \times E(G_2^2|\lambda_2) \frac{\gamma_2 + \gamma_3}{\gamma_1 + \gamma_2 + \gamma_3}
\end{aligned}$$

The weighted average power when "correct" enrollment decision is made, i.e., skip the second stage enrollment at dose step 1 under  $LFC_1^1$ , or proceed to the second stage enrollment under  $LFC_2^1$ , can be calculated as

$$\begin{aligned}
\tilde{G}_1^1 &= (\gamma_2 \text{Prob}(I_2|N_2) + \gamma_3 \text{Prob}(I_3|N_2)) / (\gamma_2 + \gamma_3) \\
\tilde{G}_2^1 &= (\gamma_1 \text{Prob}(I_1|N_1) + \gamma_2 \text{Prob}(I_2|N_1) + \gamma_3 \text{Prob}(I_3|N_1)) / (\gamma_1 + \gamma_2 + \gamma_3)
\end{aligned}$$

The optimal cut point at dose step 1  $\lambda_1$  can be derived by minimize the maximum opportunity loss as following:

$$L_1(\lambda) = \text{Max}(\tilde{G}_1^1 - E(G_1^1|\lambda, \lambda_2), \tilde{G}_2^1 - E(G_2^1|\lambda, \lambda_2))$$

In general, in a  $k$  dose group dose escalation study, the cut point at dose level  $i$  can be derived by minimizing the maximum opportunity loss:

$$L_i(\lambda) = \text{Max}(\tilde{G}_1^i - E(G_1^i|\lambda, \lambda_{i+1}, \lambda_{i+2}, \dots, \lambda_{k-1}), \tilde{G}_2^i - E(G_2^i|\lambda, \lambda_{i+1}, \lambda_{i+2}, \dots, \lambda_{k-1}))$$

The search for optimal critical values can be done by using numerical methods since the analytic methods are not available to solve the optimal cut point values. In most clinical studies where the actual response is a dichotomous variable, and the values of the cut points

are restricted to a set of limited number of integers. The numerical search of the critical values can be performed by comparing the maximum opportunity loss among all possible choices. When a continuous dose response is used, the maximum loss function about the cut points is generally unimodal, which makes the efficient numerical search feasible. However, computing algorithm to search for the continuous cut point is not pursued in this thesis.

### 4.3.3 Optimal Adaptive Design When Simultaneous Confidence Intervals are Used to Analyze the Study Data

Similar conservative stage-wise optimization approach to derive the values of the cut points for adaptive design is proposed when inference of simultaneous confidence intervals is used to analyze the study data. The least favorable configurations  $LFC_1^i$  and  $LFC_2^i$  ( $i = 1, 2, \dots, k-1$ ) defined in Section 4.3.2 have the same implication: they are the configurations under which that the probability of making the unfavorable decisions of enrollment after interim look is the biggest, while the weighted average power to identify all effective dose levels is the smallest. The proof of that the probability of making the unfavorable decisions of enrollment after interim look is the biggest is the same as I did earlier in the chapter and will not be repeated here.

Suppose after the adaptive study enrollment is finished, dose group  $k-h+1, k-h+2, \dots, k$  are fully enrolled and thus enter the final analysis, and the effective dose groups are  $k-r+1, k-r+2, \dots, k$ . In other words, there are  $r$  dose groups that are effective relative to the control group and  $h$  dose groups that enter the analysis. Let  $t$  be the maximum of  $k-r+1$  and  $k-s+1$ . Write the estimates of the drug response at the control group and each dose group as  $\bar{x}_0, \bar{x}_1, \dots, \bar{x}_k$ . Considering the max-min procedure in deriving the simultaneous confidence intervals, we have the expression of  $G$  as:

$$\begin{aligned}
G &= \sum_{i=t}^k \gamma_i Pr(I_i) / \sum_{i=t}^k \gamma_i \\
&= \sum_{i=t}^k \gamma_i Pr(\text{Reject } H_{i0}) / \sum_{i=t}^k \gamma_i \\
&= \sum_{i=t}^k \gamma_i Pr(\max_{l \leq i} \{(\bar{x}_l - \bar{x}_0) / \sqrt{\sigma^2/n_i + \sigma^2/n_0}\} > m_{\Sigma, h}^\alpha) / \sum_{i=t}^k \gamma_i
\end{aligned} \tag{4.2}$$

where  $m_{\Sigma, h}^\alpha$  is the critical value of simultaneous confidence intervals discussed in Chapter 2. The isotonic assumption of dose dose effect means and the use of max-min procedure to derive the lower bounds of SCI ensures that that  $Pr(I_k) \geq Pr(I_{k-1}) \geq \dots \geq Pr(I_{k-r+1})$ , which in turns indicates that the weighted average power to identify all effective dose levels is the smallest under  $LFC_1^i$  and  $LFC_2^i$ . Therefore the least favorable configurations  $LFC_1^i$  and  $LFC_2^i$  have the same implications when the approach of simultaneous confidence intervals is used.

In order to adjust for the adaptive procedure used in study subjects enrollment, I propose to use the following approach to derived the critical value for simultaneous confidence intervals. Suppose the common standard deviation of normally distributed drug effect variables is  $\sigma$ . At dose level  $i$  ( $i \leq k-1$ ), after the first stage enrollment  $\bar{x}_{i1} - \bar{x}_{0i} > \lambda_i$  is observed for the first time, and thus dose  $i$  and all the higher dose groups are fully enrollment, with sample size in each treatment group as  $n$ . The  $1 - \alpha$  one-sided simultaneous confidence intervals for comparisons between each dose group and the control group ( $\mu_j - \mu_0$ ,  $i \leq j \leq k$ ) can be obtained by inverting the rejection region of the following simultaneous one-sided hypothesis testing with type I error rate  $\alpha$ :

$$H_0 : \text{There is at least one } j, i \leq j \leq k, \text{ such that } \mu_j \leq \mu_0$$

$$H_1 : \mu_j > \mu_0 \text{ for all } i \leq j \leq k$$

The desired rejection region can be obtained by solving the critical value  $m$  in the following equation:

$$Prob\left(\bigcup_{i \leq j \leq k} \{\bar{x}_j - \bar{x}_0 > m_i \sigma \frac{1}{\sqrt{2/n}} \mid \mu_j = \mu_0\}\right) = \alpha$$

Take into consideration the adaptive enrollment that we use, and the positive correlations between the drug effect estimate of any dose group higher than dose  $i$ ,  $\bar{x}_h - \bar{x}_0$  ( $i \leq h \leq k$ ), and the first stage observed drug effect of any dose group lower than dose  $i$ ,  $\bar{x}_{l1} - \bar{x}_0$  ( $1 \leq l < i$ ), the following inequality can be used to find the critical values of the simultaneous confidence intervals:

$$\begin{aligned} & Prob\left(\bigcup_{i \leq j \leq k} \{\bar{x}_j - \bar{x}_0 < m_i \sigma \frac{1}{\sqrt{2/n}} \mid \mu_j = \mu_0\} \mid \bar{x}_{i1} - \bar{x}_{0i} > \lambda_i; \bar{x}_{j1} - \bar{x}_{0j} \leq \lambda_j \text{ for } 1 \leq j < i\right) \\ & \leq Prob\left(\bigcup_{i \leq j \leq k} \{\bar{x}_j - \bar{x}_0 > m_i \sigma \frac{1}{\sqrt{2/n}}\} \mid \mu_1 = \dots = \mu_j = \mu_0; \bar{x}_{i1} - \bar{x}_{0i} > \lambda_i\right) \\ & = \alpha \end{aligned}$$

The marginal multivariate distribution of observed relative drug effects  $\bar{x}_j - \bar{x}_0$  ( $i \leq j \leq k$ ) can be obtained using the Monte Carlo Composition algorithm with rejective/acceptance method. The computing algorithm introduced in chapter 2 then can be used to search for the critical value  $m$  using the empirical estimates of the correlation matrix. The max-min procedure then can be applied to obtain the simultaneous confidence intervals under order restriction  $\mu_i \leq \mu_{i+1} \leq \dots \mu_k$ :

$$\mu_j - \mu_0 \geq \max_{l \leq j} \left\{ \bar{x}_l - \bar{x}_0 - m_l \sigma \frac{1}{\sqrt{2/n}} \right\} \text{ for all } i \leq j \leq k$$

#### 4.3.4 Case Study

In this section I will illustrate the above optimal adaptive dose study designs using an example study. The example dose escalation study I am using has only three dose levels for ease of presentation. However, the design approach, analysis approaches and programming algorithm can be readily extended to dose escalation studies with more than three dose levels.

As in the previous chapters, the dose response is measured by a binary variable and we use logarithm of the odds ratio between each dose group and the control group as the study endpoint. The binary variable represents treatment success of the target disease in each study group. Write the probability of observing the treatment success within the follow up period of the study as  $p_0$  in the control group, and  $p_i$  in dose group  $i$ , where  $i = 1, 2$  or  $3$ . Suppose we know in advance that the expected success rate in the control group is 10%, and we decide that for the experiment drug to be considered as effective, the success rate needs to be increased to no less than 14%.

Write  $z_i = \log\left(\frac{p_i/(1-p_i)}{p_0/(1-p_0)}\right)$ ,  $i = 1, 2, 3$ . Dose level  $i$  will be declared to be effective if the null hypothesis  $H_{i0} : z_i \leq 0$  is rejected by one-sided test with alternative hypothesis  $H_{1i} : z_i > 0$ , and type I error probability at 0.05. The analysis approach to be used to conduct the test is the General Max-min Approach of Simultaneous Confidence Intervals we proposed in Chapter 2, or the Modified Stepdown Testing Procedure in Adaptive Dose Study with a Binary Endpoint we proposed in Chapter 3. Computer programming algorithms are developed based on the backward induction approach introduced in this Chapter.

The number of subjects at the first stage of the adaptive design in a dose escalation study is usually decided largely based on clinical considerations. Sufficient number of subjects need to be studied for the mechanism of drug effect and safety concern before the study can be



escalated to the next higher dose level. Simulation studies with different numbers of subjects at the first stage are performed to explore the design efficiency variation when the number of subjects enrolled at the first stage vary. Another parameter of such study that has multiple treatment groups and a common control group is the ratio of sample sizes between the common control and each treatment group. Williams [36] concluded that, to achieve the highest power, the ratio should be equal to approximately the square root of the number of treatment groups. However, in a phase II dose escalation study, clinicians usually prefer to have more subjects in the treatment groups to evaluate various safety aspect of the drug. Therefore, we use the same number of subjects in each dose group and the control group. The total sample size for the three dose groups and the control group is 4000.

The cut points of adaptive design are listed in the following tables. These cut points are decided in terms of the difference in numbers of treatment success between the control group and each dose group. Some of the notations in the tables are described as following:

Weights: The weights of G based on the number of effective dose levels identified (1 dose: 2 doses: 3 dose)

Dose: The dose level at which the adaptive sampling performed.

$M$ : The cut points defined in terms of the difference between the number of the observed success in dose group  $i$  ( $m_{i1}$ ) and the number of treatment success in the control group ( $m_{0i}$ ) after the first stage enrollment :  $m_{i1} - m_{0i}$

$p_1$ : The probability that the dose group has second stage enrollment under  $LFC_1$

$p_2$ : The probability that the dose group has second stage enrollment under  $LFC_2$

$G_2$ : Weighted average power under  $LFC_2$  with adaptive design

$\tilde{G}_2$ : Weighted average power under  $LFC_2$  without adaptive design

$G_1$ : Weighted average power under  $LFC_1$  with adaptive design

$\tilde{G}_1$ : Weighted average power under  $LFC_1$  without adaptive design

In each of the three tables, the benefit of using adaptive enrollment when compared to the non-adaptive enrollment with equal sample size in each treatment group is clearly seen. The weighted average power is substantially improved ( $G_1$  vs  $\tilde{G}_1$ ) when the current dose level is not effective, while the decrease of weighted average power ( $G_2$  vs  $\tilde{G}_2$ ) is almost negligible when the current dose level is actually effective. The flexibility of using appropriate weights in the weighted average power based on clinical considerations, and the easy to perform adaptive sampling rules make this design very attractive in practice.

It can also be noticed that the critical values for the adaptive design with stepdown testing procedure are consistently larger than the critical values for the adaptive design with simultaneous confidence intervals. This confirms our previous prediction that limiting enrollment in ineffective dose groups has bigger impact when the study data are analyzed using SCI, since it both increases sample sizes in effective dose groups and enables us to focus on less dose levels. Also by comparing the weighted average power across three tables, we can see that the variation of the weighted average power under both Least Favorable Configurations remain fairly stable given the same weights.

Table 11: Adaptive Dose Study Design: 200 Subjects Enrolled at the First Stage of Each Dose Level

Analysis Approach: Stepdown Testing Procedure								
Weights	Dose	$M$	$p_1$	$p_2$	$G_2$	$\tilde{G}_2$	$G_1$	$\tilde{G}_1$
1:2:3	2	1	0.467	0.877	0.850	0.869	0.958	0.917
1:2:3	1	5	0.226	0.705	0.669	0.682	0.827	0.749
1:3:5	2	0	0.533	0.906	0.837	0.846	0.958	0.917
1:3:5	1	4	0.279	0.756	0.642	0.654	0.805	0.722
1:3:7	2	0	0.533	0.906	0.838	0.846	0.958	0.917
1:3:7	1	3	0.338	0.802	0.604	0.615	0.799	0.722
1:4:9	2	0	0.533	0.906	0.827	0.840	0.958	0.917
1:4:9	1	3	0.338	0.802	0.604	0.620	0.787	0.710
Analysis Approach: Simultaneous Confidence Intervals								
Weights	Dose	$M$	$p_1$	$p_2$	$G_2$	$\tilde{G}_2$	$G_1$	$\tilde{G}_1$
1:2:3	2	1	0.467	0.877	0.882	0.900	0.908	0.832
1:2:3	1	6	0.179	0.650	0.776	0.786	0.825	0.716
1:3:5	2	0	0.533	0.906	0.859	0.877	0.908	0.832
1:3:5	1	5	0.226	0.705	0.747	0.757	0.788	0.676
1:3:7	2	0	0.533	0.905	0.859	0.877	0.908	0.832
1:3:7	1	4	0.279	0.756	0.694	0.710	0.780	0.676
1:4:9	2	0	0.533	0.905	0.844	0.865	0.908	0.832
1:4:9	1	4	0.279	0.756	0.697	0.710	0.774	0.656

Table 12: Adaptive Dose Study Design: 300 Subjects Enrolled at the First Stage of Each Dose Level

Analysis Approach: Stepdown Testing Procedure								
Weights	Dose	$M$	$p_1$	$p_2$	$G_2$	$\tilde{G}_2$	$G_1$	$\tilde{G}_1$
1:2:3	2	3	0.367	0.884	0.845	0.867	0.960	0.909
1:2:3	1	6	0.227	0.794	0.679	0.682	0.817	0.747
1:3:5	2	2	0.419	0.907	0.831	0.853	0.956	0.909
1:3:5	1	6	0.227	0.794	0.647	0.654	0.799	0.722
1:3:7	2	2	0.419	0.907	0.831	0.853	0.956	0.909
1:3:7	1	5	0.270	0.828	0.613	0.625	0.795	0.722
1:4:9	2	1	0.473	0.927	0.829	0.846	0.951	0.909
1:4:9	1	5	0.270	0.828	0.612	0.622	0.791	0.710
Analysis Approach: Simultaneous Confidence Intervals								
Weights	Dose	$M$	$p_1$	$p_2$	$G_2$	$\tilde{G}_2$	$G_1$	$\tilde{G}_1$
1:2:3	2	3	0.367	0.884	0.882	0.890	0.927	0.818
1:2:3	1	8	0.153	0.714	0.795	0.786	0.840	0.715
1:3:5	2	2	0.419	0.907	0.856	0.865	0.918	0.818
1:3:5	1	8	0.153	0.714	0.762	0.756	0.809	0.675
1:3:7	2	2	0.419	0.907	0.856	0.865	0.918	0.818
1:3:7	1	7	0.188	0.756	0.709	0.709	0.804	0.675
1:4:9	2	2	0.419	0.907	0.841	0.852	0.918	0.818
1:4:9	1	7	0.188	0.756	0.710	0.710	0.784	0.655

Table 13: Adaptive Dose Study Design: 400 Subjects Enrolled at the First Stage of Each Dose Level

Analysis Approach: Stepdown Testing Procedure								
Weights	Dose	$M$	$p_1$	$p_2$	$G_2$	$\tilde{G}_2$	$G_1$	$\tilde{G}_1$
1:2:3	2	5	0.298	0.895	0.841	0.854	0.959	0.899
1:2:3	1	9	0.158	0.793	0.681	0.682	0.821	0.747
1:3:5	2	4	0.360	0.914	0.825	0.834	0.955	0.899
1:3:5	1	8	0.188	0.823	0.653	0.654	0.800	0.722
1:3:7	2	4	0.340	0.914	0.825	0.839	0.955	0.899
1:3:7	1	7	0.222	0.850	0.635	0.797	0.625	0.722
1:4:9	2	3	0.384	0.930	0.821	0.831	0.952	0.899
1:4:9	1	7	0.222	0.850	0.619	0.622	0.785	0.710
Analysis Approach: Simultaneous Confidence Intervals								
Weights	Dose	$M$	$p_1$	$p_2$	$G_2$	$\tilde{G}_2$	$G_1$	$\tilde{G}_1$
1:2:3	2	5	0.297	0.895	0.880	0.889	0.930	0.803
1:2:3	1	11	0.107	0.725	0.803	0.806	0.840	0.715
1:3:5	2	4	0.339	0.913	0.851	0.853	0.923	0.803
1:3:5	1	10	0.131	0.760	0.772	0.786	0.803	0.675
1:3:7	2	4	0.339	0.913	0.851	0.853	0.923	0.803
1:3:7	1	9	0.157	0.793	0.719	0.724	0.799	0.675
1:4:9	2	3	0.383	0.929	0.840	0.839	0.915	0.803
1:4:9	1	9	0.157	0.793	0.722	0.722	0.789	0.655

## 5.0 FUTURE RESEARCH PLAN

Although the General Max-Min procedure for deriving the simultaneous confidence intervals is developed under the assumption of multivariate normal distribution, it is certainly feasible to extend the approach to other multivariate distribution. Specifically, in a clinical study with adaptive enrollment design, the distribution of the study data is not normal even if the original drug responses are normally distributed. I plan to extend the computing approach based on Somerville's algorithm to simulate the critical values of the simultaneous confidence intervals with the biased sample from adaptive clinical study.

Schoenfeld [22] (1986) suggested that, under the assumption of isotonic dose response, one can use the isotonic regression estimates and likelihood ratio test to derive the critical point for one-sided test in dose study. The test can improve the efficiency over the usual test performed individually at each dose level without considering the monotonicity of the dose response means. When the dose response means at the tested dose levels are very close to each other, which is usually the case in well-studied disease areas, Schoenfeld's test is especially powerful. Therefore, by adopting Schoenfeld's test at each step of a stepdown testing procedure, we can increase the power the testing procedure. However, Schoenfeld's approach is limited to the cases of independent dose responses. In fact, most of the isotonic regression related research is limited to the independent cases. I plan to develop a testing procedure similar to Schoenfeld's for correlated dose response, which also utilizes the isotonic

dose response assumption to improves the power of the test.

As we have illustrated in the previous chapter, the optimal parameters of the adaptive design depend on the approach used to analyze the study data. Once we have improved the analysis methods, the adaptive design parameters should be updated, too. Also I plan to extend the design approach to the more general situations when the dose study can be restricted by different clinical and logistic constraints. Also I plan to explore the adaptive design of dose study when the study endpoint is an ordinal and/or Poisson random variable.

## APPENDIX A

### SAS PROGRAMS USED TO DERIVE THE CRITICAL VALUES OF MODIFIED SIMULTANEOUS CONFIDENCE INTERVALS

/\* Input parameter for Somerville Simulation Programs

SAS/IML program for calculating critical values of multiple  
comparison procedures.

Input : (Required parameters)

swcov : Flag to include var-covar matrix (=0) or not (=1)

swtype : Type of multiple comparison procedure

conf : Confidence level

givense : User requested SE of estimates, if alternate  
procedure is used

k : Number of treatment groups

seed : Seed for random number generator

ndenom : Degrees of freedom

mocar : Number of random directions

irep : Number of preliminary estimates (if negative,  
alternate procedure is performed)

(Optional parameters)

mm : Number of contrasts (if swtype = 0)

vc : Var-covar matrix (if swcov = 0)

contr : Contrasts (if swtype = 0 or 20)

icontrol: Control group {if [(swtype=4 OR 5) AND (swcov=0)]  
OR [(swtype=9)]}

iflo : Lower bound for peak (swtype = 10)

ighi : Higher bound for peak (swtype = 10)



```

        znn      : Vector of sample sizes (swtype = 11-15)
Output :
*/

** Derive critical values when difference of probailitie is used **;
%macro rn(nn1=, nn2=, nn3=, nn4=, nn5=, nn6=, nn7=);

%do ii=1 %to 2;
    %do iii=1 %to 7;
        %let n&iii=%scan(&&nn&iii,&ii);
    %end;

data _null_;
    array varlist v1-v14;
    do i=4 to 17;
        varlist(i-3)=(i/100)*(1-i/100);
    end;
    call symput('var4', trim(left(put(v1, 8.6))));
    call symput('var5', trim(left(put(v2, 8.6))));
    call symput('var6', trim(left(put(v3, 8.6))));
    call symput('var7', trim(left(put(v4, 8.6))));
    call symput('var8', trim(left(put(v5, 8.6))));
    call symput('var9', trim(left(put(v6, 8.6))));
    call symput('var10', trim(left(put(v7, 8.6))));
    call symput('var11', trim(left(put(v8, 8.6))));
    call symput('var12', trim(left(put(v9, 8.6))));
    call symput('var13', trim(left(put(v10, 8.6))));
    call symput('var14', trim(left(put(v11, 8.6))));
    call symput('var15', trim(left(put(v12, 8.6))));
    call symput('var16', trim(left(put(v13, 8.6))));
    call symput('var17', trim(left(put(v14, 8.6))));
run;

%let charc1=" &var10 &var10 &var10 &var10 &var10 &var10 &var10";
%let charc2=" &var10 &var12 &var12 &var12 &var12 &var12 &var12";
%let charc3=" &var10 &var8 &var8 &var8 &var8 &var8 &var8 ";
%let charc4=" &var10 &var12 &var11 &var10 &var9 &var8 &var7 ";

```

```

%let charc5=" &var10 &var12 &var12 &var10 &var10 &var8 &var8 ";
%let charc6=" &var10 &var12 &var12 &var12 &var8 &var8 &var8 ";
%let charc7=" &var10 &var9 &var8 &var7 &var6 &var6 &var4 ";
%let charc8=" &var10 &var17 &var16 &var15 &var14 &var13 &var12";
%let charc9=" &var10 &var12 &var8 &var8 &var8 &var8 &var8 ";
%let charc10="&var10 &var8 &var8 &var8 &var8 &var8 &var12";

```

```

%do j=1 %to 10;

```

```

data _null_;

```

```

    array varlist v1-v7;

```

```

    %do i=1 %to 7;

```

```

        varlist(&i)=%scan(&&charc&j,&i," " )/&&n&i;

```

```

    %end;

```

```

    call symput('va1', trim(left(put(v1, 8.6))));

```

```

    call symput('va2', trim(left(put(v2, 8.6))));

```

```

    call symput('va3', trim(left(put(v3, 8.6))));

```

```

    call symput('va4', trim(left(put(v4, 8.6))));

```

```

    call symput('va5', trim(left(put(v5, 8.6))));

```

```

    call symput('va6', trim(left(put(v6, 8.6))));

```

```

    call symput('va7', trim(left(put(v7, 8.6))));

```

```

run;

```

```

options nosource;

```

```

proc iml;

```

```

    /* INSERT SITUATION PARAMETERS HERE */

```

```

    swcov = 0 ;

```

```

    swtype = 0 ;

```

```

    conf = .975 ;

```

```

    givense= 999 ;

```

```

    k = 7 ;

```

```

    seed = 1893 ;

```

```

    ndenom = -1 ;

```

```

mocar = 5000 ;
irep = 10 ;
mm = 6 ;

vc={&va1      .      .      .      .      .      .      ,
    0      &va2      .      .      .      .      .      ,
    0      0      &va3      .      .      .      .      ,
    0      0      0      &va4      .      .      .      ,
    0      0      0      0      &va5      .      .      ,
    0      0      0      0      0      &va6      .      ,
    0      0      0      0      0      0      &va7 } ;

contr={1  0  0  0  0  0 -1,
       1  0  0  0  0 -1  0,
       1  0  0  0 -1  0  0,
       1  0  0 -1  0  0  0,
       1  0 -1  0  0  0  0,
       1 -1  0  0  0  0  0};

/* END SITUATION PARAMETERS */

.....

(simulation programs by Somerville)

%end;
%end;
*quit;

%mend;

%rn(nn1=1500 1500, nn2=500 500, nn3=500 500, nn4=1000 500, nn5=1000 500,
    nn6=1000 1000, nn7=1000 1000)

** Derive critical values when relative risk is used **;

%macro rn(nn1=, nn2=, nn3=, nn4=, nn5=, nn6=, nn7=);

```

```

%do ii=1 %to 2;
  %do iii=1 %to 7;
    %let n&iii=%scan(&&nn&iii,&ii);
  %end;

data _null_;
  array varlist v1-v17;
  do i=4 to 17;
    varlist(i-3)=(1-i/100);
  end;
  call symput('var4', trim(left(put(v1, 8.6))));
  call symput('var5', trim(left(put(v2, 8.6))));
  call symput('var6', trim(left(put(v3, 8.6))));
  call symput('var7', trim(left(put(v4, 8.6))));
  call symput('var8', trim(left(put(v5, 8.6))));
  call symput('var9', trim(left(put(v6, 8.6))));
  call symput('var10', trim(left(put(v7, 8.6))));
  call symput('var11', trim(left(put(v8, 8.6))));
  call symput('var12', trim(left(put(v9, 8.6))));
  call symput('var13', trim(left(put(v10, 8.6))));
  call symput('var14', trim(left(put(v11, 8.6))));
  call symput('var15', trim(left(put(v12, 8.6))));
  call symput('var16', trim(left(put(v13, 8.6))));
  call symput('var17', trim(left(put(v14, 8.6))));
run;

%let charc1="&var10 &var10 &var10 &var10 &var10 &var10 &var10";
%let charc2="&var10 &var12 &var12 &var12 &var12 &var12 &var12 &var12";
%let charc3="&var10 &var8 &var8 &var8 &var8 &var8 &var8";
%let charc4="&var10 &var12 &var11 &var10 &var9 &var8 &var7";
%let charc5="&var10 &var12 &var12 &var10 &var10 &var8 &var8";
%let charc6="&var10 &var12 &var12 &var12 &var8 &var8 &var8";
%let charc7="&var10 &var9 &var8 &var7 &var6 &var6 &var4";
%let charc8="&var10 &var17 &var16 &var15 &var14 &var13 &var12";
%let charc9="&var10 &var12 &var8 &var8 &var8 &var8 &var8";
%let charc10="&var10 &var8 &var8 &var8 &var8 &var8 &var12";

```

```

%do j=1 %to 10;

options nosource;

data _null_;
  array varlist v1-v7;
  %do i=1 %to 7;
    p_1=%scan(&&charc&j, &i, " ");
    varlist(&i)=p_1/(&&n&i*(1-p_1));
  %end;
  call symput('va1', trim(left(put(v1, 8.6))));
  call symput('va2', trim(left(put(v2, 8.6))));
  call symput('va3', trim(left(put(v3, 8.6))));
  call symput('va4', trim(left(put(v4, 8.6))));
  call symput('va5', trim(left(put(v5, 8.6))));
  call symput('va6', trim(left(put(v6, 8.6))));
  call symput('va7', trim(left(put(v7, 8.6))));
run;

proc iml;

  /* INSERT SITUATION PARAMETERS HERE */

  swcov = 0      ;
  swtype = 0     ;
  conf   = .975  ;
  givense= 999   ;
  k      = 7     ;
  seed   = 1893  ;
  ndenom = -1    ;
  mocar  = 5000  ;
  irep   = 10    ;
  mm     = 6     ;

  vc={&va1      .      .      .      .      .      .      .      ,
      0         &va2   .      .      .      .      .      .      ,
      0         0      &va3  .      .      .      .      .      ,

```

```

0          0          0          &va4          .          .          .          ,
0          0          0          0          &va5          .          .          ,
0          0          0          0          0          &va6          .          ,
0          0          0          0          0          0          0          &va7 } ;

contr={1  0  0  0  0  0 -1,
       1  0  0  0  0 -1  0,
       1  0  0  0 -1  0  0,
       1  0  0 -1  0  0  0,
       1  0 -1  0  0  0  0,
       1 -1  0  0  0  0  0};

/* END SITUATION PARAMETERS */
.....

(simulation programs by Somerville)

%end;
%end;
*quit;

%mend;

%rn(nn1=1500 1500, nn2=500 500, nn3=500 500, nn4=1000 500, nn5=1000 500,
    nn6=1000 1000, nn7=1000 1000)

** Derive critical values when log odds ratio is used **;

%macro rn(nn1=, nn2=, nn3=, nn4=, nn5=, nn6=, nn7=);

%do ii=1 %to 2;
    %do iii=1 %to 7;
        %let n&iii=%scan(&&nn&iii,&ii);
    %end;
%end;

data _null_;
    array varlist v1-v17;
    do i=4 to 17;

```

```

    varlist(i-3)=1/(i/100)+1/(1-i/100);
end;
call symput('var4', trim(left(put(v1, 8.6))));
call symput('var5', trim(left(put(v2, 8.6))));
call symput('var6', trim(left(put(v3, 8.6))));
call symput('var7', trim(left(put(v4, 8.6))));
call symput('var8', trim(left(put(v5, 8.6))));
call symput('var9', trim(left(put(v6, 8.6))));
call symput('var10', trim(left(put(v7, 8.6))));
call symput('var11', trim(left(put(v8, 8.6))));
call symput('var12', trim(left(put(v9, 8.6))));
call symput('var13', trim(left(put(v10, 8.6))));
call symput('var14', trim(left(put(v11, 8.6))));
call symput('var15', trim(left(put(v12, 8.6))));
call symput('var16', trim(left(put(v13, 8.6))));
call symput('var17', trim(left(put(v14, 8.6))));
run;

%let charc1="&var10 &var10 &var10 &var10 &var10 &var10 &var10";
%let charc2="&var10 &var12 &var12 &var12 &var12 &var12 &var12 &var12";
%let charc3="&var10 &var8 &var8 &var8 &var8 &var8 &var8";
%let charc4="&var10 &var12 &var11 &var10 &var9 &var8 &var7";
%let charc5="&var10 &var12 &var12 &var10 &var10 &var8 &var8";
%let charc6="&var10 &var12 &var12 &var12 &var8 &var8 &var8";
%let charc7="&var10 &var9 &var8 &var7 &var6 &var6 &var4";
%let charc8="&var10 &var17 &var16 &var15 &var14 &var13 &var12";
%let charc9="&var10 &var12 &var8 &var8 &var8 &var8 &var8";
%let charc10="&var10 &var8 &var8 &var8 &var8 &var8 &var12";

%do j=1 %to 10;

options nosource;

data _null_;
    array varlist v1-v7;
    %do i=1 %to 7;
        varlist(&i)=%scan(&&charc&j, &i, " ")/&&n&i;
    %end;

```

```

%end;
call symput('va1', trim(left(put(v1, 8.6))));
call symput('va2', trim(left(put(v2, 8.6))));
call symput('va3', trim(left(put(v3, 8.6))));
call symput('va4', trim(left(put(v4, 8.6))));
call symput('va5', trim(left(put(v5, 8.6))));
call symput('va6', trim(left(put(v6, 8.6))));
call symput('va7', trim(left(put(v7, 8.6))));
run;

proc iml;

/* INSERT SITUATION PARAMETERS HERE */

swcov = 0 ;
swtype = 0 ;
conf = .975 ;
givense= 999 ;
k = 7 ;
seed = 1893 ;
ndenom = -1 ;
mocar = 5000 ;
irep = 10 ;
mm = 6 ;

vc={&va1 . . . . . ,
    0 &va2 . . . . . ,
    0 0 &va3 . . . . . ,
    0 0 0 &va4 . . . . . ,
    0 0 0 0 &va5 . . . . . ,
    0 0 0 0 0 &va6 . . . . . ,
    0 0 0 0 0 0 &va7 } ;

contr={1 0 0 0 0 0 -1,
        1 0 0 0 0 -1 0,
        1 0 0 0 -1 0 0,
        1 0 0 -1 0 0 0,
        1 0 -1 0 0 0 0,

```



```

        1 -1  0  0  0  0  0};

/* END SITUATION PARAMETERS */
%end;
%end;
*quit;

%mend;

%rn(nn1=1500 1500, nn2=500 500, nn3=500 500, nn4=1000 500, nn5=1000 500,
    nn6=1000 1000, nn7=1000 1000)

** Derive critical values when difference of arcsinc of square root of
    probailitie is used **;

%macro rn(nn1=, nn2=, nn3=, nn4=, nn5=, nn6=, nn7=);

%do ii=1 %to 2;
    %do iii=1 %to 7;
        %let n&iii=%scan(&&nn&iii,&ii);
    %end;

data _null_;
    array varlist v1-v14;
    do i=4 to 17;
        varlist(i-3)=0.25;
    end;
    call symput('var4', trim(left(put(v1, 8.6))));
    call symput('var5', trim(left(put(v2, 8.6))));
    call symput('var6', trim(left(put(v3, 8.6))));
    call symput('var7', trim(left(put(v4, 8.6))));
    call symput('var8', trim(left(put(v5, 8.6))));
    call symput('var9', trim(left(put(v6, 8.6))));
    call symput('var10', trim(left(put(v7, 8.6))));
    call symput('var11', trim(left(put(v8, 8.6))));
    call symput('var12', trim(left(put(v9, 8.6))));
    call symput('var13', trim(left(put(v10, 8.6))));

```

```

call symput('var14', trim(left(put(v11, 8.6))));
call symput('var15', trim(left(put(v12, 8.6))));
call symput('var16', trim(left(put(v13, 8.6))));
call symput('var17', trim(left(put(v14, 8.6))));
run;

%let charc1="&var10 &var10 &var10 &var10 &var10 &var10 &var10";
%let charc2="&var10 &var12 &var12 &var12 &var12 &var12 &var12 &var12";
%let charc3="&var10 &var8 &var8 &var8 &var8 &var8 &var8";
%let charc4="&var10 &var12 &var11 &var10 &var9 &var8 &var7";
%let charc5="&var10 &var12 &var12 &var10 &var10 &var8 &var8";
%let charc6="&var10 &var12 &var12 &var12 &var8 &var8 &var8";
%let charc7="&var10 &var9 &var8 &var7 &var6 &var6 &var4";
%let charc8="&var10 &var17 &var16 &var15 &var14 &var13 &var12";
%let charc9="&var10 &var12 &var8 &var8 &var8 &var8 &var8";
%let charc10="&var10 &var8 &var8 &var8 &var8 &var8 &var12";

%do j=1 %to 10;

options nosource;

data _null_;
  array varlist v1-v7;
  %do i=1 %to 7;
    varlist(&i)=%scan(&&charc&j,&i," ")/&&n&i;
  %end;
  call symput('va1', trim(left(put(v1, 8.6))));
  call symput('va2', trim(left(put(v2, 8.6))));
  call symput('va3', trim(left(put(v3, 8.6))));
  call symput('va4', trim(left(put(v4, 8.6))));
  call symput('va5', trim(left(put(v5, 8.6))));
  call symput('va6', trim(left(put(v6, 8.6))));
  call symput('va7', trim(left(put(v7, 8.6))));
run;

proc iml;

```

```
/* INSERT SITUATION PARAMETERS HERE */
```

```
swcov = 0 ;  
swtype = 0 ;  
conf = .975 ;  
givense= 999 ;  
k = 7 ;  
seed = 1893 ;  
ndenom = -1 ;  
mocar = 5000 ;  
irep = 10 ;  
mm = 6 ;
```

```
vc={&va1 . . . . . ,  
0 &va2 . . . . . ,  
0 0 &va3 . . . . . ,  
0 0 0 &va4 . . . . . ,  
0 0 0 0 &va5 . . . . . ,  
0 0 0 0 0 &va6 . . . . . ,  
0 0 0 0 0 0 &va7 } ;
```

```
contr={1 0 0 0 0 0 -1,  
1 0 0 0 0 -1 0,  
1 0 0 0 -1 0 0,  
1 0 0 -1 0 0 0,  
1 0 -1 0 0 0 0,  
1 -1 0 0 0 0 0};
```

```
/* END SITUATION PARAMETERS */
```

```
%end;  
%end;  
*quit;
```

```
%mend;
```

```
%rn(nn1=1500 1500, nn2=500 500, nn3=500 500, nn4=1000 500, nn5=1000 500,
```

```

nn6=1000 1000, nn7=1000 1000)

** Programs to obtain max-min Simultaneous Confidence Intervals for
simulated data **;

libname save '/proj/c5/combo/gurm/wbc_10107/analysis';
libname dat '/proj/c5/combo/gurm/wbc_10107/analysis/ana/proposal/programs';

%macro simu(dat=,l=);

proc sort data=&dat out=simu; by situ;
run;

data ana6;
merge simu
      dat.crit_multinorm;
by situ;
array crt{4}    crit_p crit_ars crit_lod crit_lrr;
array low{4,6}  low_p1-low_p6 low_ars1-low_ars6 low_lod1-low_lod6
                low_lrr1-low_lrr6;
array up{4,6}   up_p1-up_p6  up_ars1-up_ars6  up_lod1-up_lod6
                up_lrr1-up_lrr6;
array wid{4,6}  widp1-widp6  widars1-widars6  widlod1-widlod6
                widlrr1-widlrr6;
array wid1{4,6} wid1p1-wid1p6 wid1ars1-wid1ars6 wid1lod1-wid1lod6
                wid1lrr1-wid1lrr6;
array ral{4,6}  dp_r1-dp_r6 dap_r1-dap_r6 lods_r1-lods_r6 lrr_r1-lrr_r6;
array indc{4}   dp_i dap_i lods_i lrr_i;
array m{4,6}    dp1-dp6 dap1-dap6 lods1-lods6 lrr1-lrr6;
array q{4,6}    dpq1-dpq6 dapq1-dapq6 lodsq1-lodsq6 lrrq1-lrrq6;
do j=1 to 4;
    indc(j)=1;
    do h=1 to 6;
        low(j,h)=ral(j,h)-q(j,h)*crt(j);
        up(j,h)=ral(j,h)+q(j,h)*crt(j);
        wid1(j,h)=up(j,h)-low(j,h);
        do ii=1 to h;

```

```

        up(j,h)=min(ral(j,ii)+q(j,ii)*crt(j),up(j,h));
    end;
    do iii=h to 6;
        low(j,h)=max(ral(j,iii)-q(j,iii)*crt(j),low(j,h));
    end;
    if m(j,h)>ral(j,h)+q(j,h)*crt(j) or m(j,h)<ral(j,h)
        -q(j,h)*crt(j) then indc(j)=0;
    wid(j,h)=up(j,h)-low(j,h);
end;

end;

widavp=mean(widp1, widp2, widp3, widp4, widp5, widp6);
widavars=mean(widars1, widars2, widars3, widars4, widars5, widars6);
widavlod=mean(widlod1, widlod2, widlod3, widlod4, widlod5, widlod6);
widavlrr=mean(widlrr1, widlrr2, widlrr3, widlrr4, widlrr5, widlrr6);
widav1p=mean(wid1p1, wid1p2, wid1p3, wid1p4, wid1p5, wid1p6);
widav1ars=mean(wid1ars1,wid1ars2,wid1ars3,wid1ars4,wid1ars5,wid1ars6);
widav1lod=mean(wid1lod1,wid1lod2,wid1lod3,wid1lod4,wid1lod5,wid1lod6);
widav1lrr=mean(wid1lrr1,wid1lrr2,wid1lrr3,wid1lrr4,wid1lrr5,wid1lrr6);
ratio_p=mean(widp1/wid1p1,widp2/wid1p2,widp3/wid1p3,widp4/wid1p4,
            widp5/wid1p5,widp6/wid1p6);
ratio_ars=mean(widars1/wid1ars1,widars2/wid1ars2,widars3/wid1ars3,
            widars4/wid1ars4,widars5/wid1ars5,widars6/wid1ars6);
ratio_lod=mean(widlod1/wid1lod1,widlod2/wid1lod2,widlod3/wid1lod3,
            widlod4/wid1lod4,widlod5/wid1lod5,widlod6/wid1lod6);
ratio_lrr=mean(widlrr1/wid1lrr1,widlrr2/wid1lrr2,widlrr3/wid1lrr3,
            widlrr4/wid1lrr4,widlrr5/wid1lrr5,widlrr6/wid1lrr6);

run;

title "&l";

proc means data=ana6 noprint;
    var dp_i dap_i lods_i lrr_i;
    class situ;
    output out=out;
run;

data out1;

```

```

set out;
where _stat_="MEAN";
keep situ dp_i dap_i lods_i lrr_i;
run;

proc print; run;

/*
proc means data=ana6 noprint;
var widavp widav1p ratio_p widavars widav1ars ratio_ars widavlod
    widav1lod ratio_lod widavlrr widav1lrr ratio_lrr;
class situ;
output out=out;
run;

data out1;
set out;
where _stat_="MEAN";
keep situ widavp widav1p ratio_p widavars widav1ars ratio_ars
    widavlod widav1lod ratio_lod widavlrr widav1lrr ratio_lrr;
run;

proc print; run;
*/
%mend;

%simu(dat=save.simu10000_uq, l='control-30000 (123)-10000 (456)-20000');

```

## APPENDIX B

### SAS PROGRAMS USED TO DERIVED THE CRITICAL VALUES FOR MIXTURE NORMAL APPROXIMATION APPROACH

```
options symbolgen mprint mlogic;

%macro monte(p=,pa=,ni1=,n01=,n0=,ni=,d=, msz=);

title "P=&p Pa=&pa Sample Size(TRT:Control): N1=&ni1:&n01 N=&ni:&n0";

data report;
  dm=.; pct_low=.; err90=.; power90=.; pct_med=.; err95=.;
  power95=.; pct_high=.; err975=.; power975=.; errr90=.; powerr90=.;
  errr95=.; powerr95=.; errr975=.; powerr975=.;
run;

%let i=1;
%do %until(%scan(&d, &i, " ")=);
  %let di=%scan(&d, &i);
  %if %substr(&di,1,1)=n %then %do;
    %let di=(-%substr(&di,2));
  %end;

data tmp1;
  d=&di;
  p1=0;
  ** probability of crossing cut point after first stage under
  null hypothesis **;
```

```

do m=0 to &n01;
  do n=0 to &ni1;
    if n-m>=&di then do;
      if m>0 and n>0 then tmp=(probbnml(&p,&n01,m)-probbnml(&p,&n01,m-1))
        *(probbnml(&p,&ni1,n)-probbnml(&p,&ni1,n-1));
      if m=0 and n>0 then tmp=probbnml(&p,&n01,m)
        *(probbnml(&p,&ni1,n)-probbnml(&p,&ni1,n-1));
      if m>0 and n=0 then tmp=(probbnml(&p,&n01,m)-probbnml(&p,&n01,m-1))
        *probbnml(&p,&ni1,n);
      if m=0 and n=0 then tmp=probbnml(&p,&n01,m)*probbnml(&p,&ni1,n);
    end;
    else tmp=0;
  p1=p1+tmp;
  end;
end;
p90=trim(left(put((1-0.1/p1)*100,8.7)));
p95=trim(left(put((1-0.05/p1)*100,8.7)));
p975=trim(left(put((1-0.025/p1)*100,8.7)));
call symput("p90",p90);
call symput("p95",p95);
call symput("p975",p975);
p2=0;
** probability of crossing cut point after first stage under alternative
  hypothesis **;
do m=0 to &n01;
  do n=0 to &ni1;
    if n-m>=&di then do;
      if m>0 and n>0 then tmp=(probbnml(&p,&n01,m)-probbnml(&p,&n01,m-1))
        *(probbnml(&pa,&ni1,n)-probbnml(&pa,&ni1,n-1));
      if m=0 and n>0 then tmp=probbnml(&p,&n01,m)
        *(probbnml(&pa,&ni1,n)-probbnml(&pa,&ni1,n-1));
      if m>0 and n=0 then tmp=(probbnml(&p,&n01,m)-probbnml(&p,&n01,m-1))
        *probbnml(&pa,&ni1,n);
      if m=0 and n=0 then tmp=probbnml(&p,&n01,m)*probbnml(&pa,&ni1,n);
    end;
    else tmp=0;
  p2=p2+tmp;

```



```

        end;
    end;
    p_1=trim(left(put(p1,8.7)));
    p_2=trim(left(put(p2,8.7)));
    call symput("p1",p_1);
    call symput("p2",p_2);
run;
/*
proc print; var d p1 p2 p90 p95 p975; run;
*/
data tmp2;
    s1=sqrt(1/(&p*&n01)+1/(&n01-&p*&n01)+1/(&p*&ni1)+1/(&ni1-&p*&ni1));
    s2=sqrt(1/(&p*&n0)+1/(&n0-&p*&n0)+1/(&p*&ni)+1/(&ni-&p*&ni));
    rho=(s2*s2)/(s1*s2);
    pp95=trim(left(put(s2*probit(0.95),8.7)));
    pp90=trim(left(put(s2*probit(0.9),8.7)));
    pp975=trim(left(put(s2*probit(0.975),8.7)));
    call symput("pp90",pp90);
    call symput("pp95",pp95);
    call symput("pp975",pp975);
    ** mixture normal distribution simulation **;
    do i=1 to &msz;
        m=ranbin(1009, &n01, &p);
        if m+&di>0 then do;
            n=m+&di-.5;
            lambda=((n+0.5)/(&ni1-n+0.5)) / ((m+.5)/(&n01-m+.5));
        end;
        else if m+&di<=0 then lambda=0.00000001;
        k=0;
        do until(k>10000);
            z1=rannor(k);
            z11=z1*s1;
            if z11>=log(lambda) then k=10001;
        end;
        z2=rannor(i);
        zi=rho*s2*z1+sqrt(1-rho*rho)*s2*z2;
        output;
    end;

```

```

end;
** exact distribution simulation **;
n02=&n0-&n01;
ni2=&ni-&ni1;
i=0;
do until(i>=&msz);
  m01=ranbin(1009,&n01,&p);
  mi1=ranbin(1009,&ni1,&p);
  if mi1-m01>=&di then do;
    m02=ranbin(1009,n02,&p);
    mi2=ranbin(1009,ni2,&p);
    m0=m01+m02;
    mi=mi1+mi2;
    odds=((mi+0.5)/(&ni-mi+0.5))/((m0+0.5)/(&n0-m0+0.5));
    zi=log(odds);
    output;
    i=i+1;
  end;
end;*/
run;

** percentiles of the marginal distribution of testing statistics **;
proc univariate data=tmp2 noprint;
  var zi;
  output out=out mean=mean std=std pctlpts=&p90 &p95 &p975 pctlpre=pct_
    pctlname=low med high ;
run;

/*
proc print data=out; run;
*/

data validate dat;
  set out;
  n02=&n0-&n01;
  ni2=&ni-&ni1;
  ** error rates when adjusted critical values used **;

```

```

err90=0;
err95=0;
err975=0;
** error rates when unadjusted critical values used **;
errr90=0;
errr95=0;
errr975=0;
** simulate the error rate **;
do i=1 to &msz;
  m01=ranbin(1009,&n01,&p);
  mi1=ranbin(1009,&ni1,&p);
  if mi1-m01>=&di then do;
    m02=ranbin(1009,n02,&p);
    mi2=ranbin(1009,ni2,&p);
    m0=m01+m02;
    mi=mi1+mi2;
    odds=((mi+0.5)/(&ni-mi+0.5))/((m0+0.5)/(&n0-m0+0.5));
    z=log(odds);
    output dat;
    if z>pct_low then err90=err90+1;
    if z>pct_med then err95=err95+1;
    if z>pct_high then err975=err975+1;
    if z>&pp90 then errr90=errr90+1;
    if z>&pp95 then errr95=errr95+1;
    if z>&pp975 then errr975=errr975+1;
  end;
end;
err90=err90/&msz;
err95=err95/&msz;
err975=err975/&msz;
errr90=errr90/&msz;
errr95=errr95/&msz;
errr975=errr975/&msz;
** testing power when adjusted critical values used **;
power90=0;
power95=0;
power975=0;

```

```

** testing power when unadjusted critical values used **;
powerr90=0;
powerr95=0;
powerr975=0;
do j=1 to &msz;
  m01=ranbin(1009,&n01,&p) ;
  mi1=ranbin(1009,&ni1,&pa);
  if mi1-m01>=&di then do;
    m02=ranbin(1009,&n0.-&n01,&p);
    mi2=ranbin(1009,&ni.-&ni1,&pa);
    m0=m01+m02;
    mi=mi1+mi2;
    odds=((mi+0.5)/(&ni-mi+0.5))/((m0+0.5)/(&n0-m0+0.5));
    z=log(odds);
    if z>pct_low then power90=power90+1;
    if z>pct_med then power95=power95+1;
    if z>pct_high then power975=power975+1;
    if z>&pp90 then powerr90=powerr90+1;
    if z>&pp95 then powerr95=powerr95+1;
    if z>&pp975 then powerr975=powerr975+1;
  end;
end;
power90=power90/&msz;
power95=power95/&msz;
power975=power975/&msz;
powerr90=powerr90/&msz;
powerr95=powerr95/&msz;
powerr975=powerr975/&msz;
dm=&di+0;
p1=&p1+0;
p2=&p2+0;
output validate;
keep z dm pct_low err90 power90 pct_med err95 power95 pct_high err975
  power975 errr90 powerr90 errr95 powerr95 errr975 powerr975 p1 p2;
run;
/*
proc univariate data=dat;

```

```

    var z;
run;
*/
data report;
    set report validate;
run;

%let i=%eval(&i+1); %end;

data report;
    set report;
    if _n_=1 then delete;
run;

proc print noobs;
    var dm p1 pct_low  err90  power90  errr90  powerr90;
run;
proc print noobs;
    var dm p1 pct_med  err95  power95  errr95  powerr95;
run;
proc print noobs;
    var dm p1 pct_high err975 power975 errr975 powerr975;
run;

%mend;

%monte(p=0.1, pa=0.15, ni1=150, n01=150, n0=500, ni=500,
    msz=200000, d=n8 n7 n6 n5 n4 n3 n2 n1 0 1 2 3 4 5);

%monte(p=0.1, pa=0.13, ni1=300, n01=300, n0=1000, ni=1000, msz=200000,
d=n15 n14 n13 n12 n11 n10 n9 n8 n7 n6 n5 n4 n3 n2 n1 0 1 2 3 4 5 6 7 8 9);

```

## APPENDIX C

### SAS PROGRAMS USED TO OBTAIN THE OPTIMAL CUT POINTS WHEN SCI APPROACHED IS USED

```
options symbolgen mprint mlogic;
%macro dsn(p=, pa=, ni1=, n01=, nni=, nn0=, msz=, r3=, r2=, r1=, seed=1009);

%let cr3=1.645;
%let cr2=2.080;
%let cr1=2.329;

data design;
  dose=.; di=.; p1=.; p2=.; pct_pt2=.; pct_pt3=.; err2=.; err3=.; power2=.;
  power3=.; g11=.; g01=.; g03=.; l1=.; g01=.; g02=.; l0=.; l_max=.; g_min=.;
run;

%do i=1 %to 2;

data report;
  dose=.; di=.; p1=.; p2=.; pct_pt2=.; pct_pt3=.; err2=.; err3=.; power2=.;
  power3=.; g11=.; g01=.; g03=.; l1=.; g01=.; g02=.; l0=.; l_max=.; g_min=.;
run;

%let dose=%eval(3-&i);
%let pdose=%eval(3-&i+1);

** sample sizes at dose step 1 and first stage of dose step 2 **;
%let ni=%sysevalf((&nni-&ni1)*(&dose-1)/(3-&dose+1)+&nni);
```

```

%let n0=%sysevalf((&nn0-&n01)*(&dose-1)/(3-&dose+1)+&nn0);
%let ni2=%sysevalf((&nni-&ni1)*(&dose)/(3-&dose)+&nni);
%let n02=%sysevalf((&nn0-&n01)*(&dose)/(3-&dose)+&nn0);

data _null_;
  ** standard error of log odds ratio at dose level 1 & 2 **;
  mu=log(&pa/(1-&pa)/(&p/(1-&p)));
  s1=sqrt(1/(&p*&n01)+1/(&n01-&p*&n01)+1/(&p*&ni1)+1/(&ni1-&p*&ni1));
  s2=sqrt(1/(&p*&n0)+1/(&n0-&p*&n0)+1/(&p*&ni)+1/(&ni-&p*&ni));
  s3=sqrt(1/(&p*&n02)+1/(&n02-&p*&n02)+1/(&p*&ni2)+1/(&ni2-&p*&ni2));
  rho=(s2*s2)/(s1*s2);
  call symput("s1", trim(left(put(s1,10.6))));
  call symput("s2", trim(left(put(s2,10.6))));
  call symput("s3", trim(left(put(s3,10.6))));
  call symput("rho", trim(left(put(rho,10.6))));
  call symput("mu", trim(left(put(mu,10.6))));
run;

data rn1 rn2 rn3;
  ** simulated sample to be used in the program **;
  do i=1 to &msz;
    m=ranbin(&seed,&n01,&p);
    z1=rannor(&seed);
    z2=rannor(&seed);
    output rn1;
  end;
  do i=1 to &msz;
    m02=ranbin(&seed,&n0.-&n01,&p);
    mi2=ranbin(&seed,&ni.-&ni1,&p);
    m01=ranbin(&seed,&n01,&p) ;
    mi1=ranbin(&seed,&ni1,&p);
    output rn2;
  end;
  do i=1 to &msz;
    m01=ranbin(&seed,&n01,&p);
    mi1=ranbin(&seed,&ni1,&pa);
    m02=ranbin(&seed,&n0.-&n01,&p);

```

```

        mi2=ranbin(&seed,&ni.-&ni1,&pa);
        output rn3;
    end;
run;

%do di=-50 %to 50;

%let va=0;

data _null_;
    p1=0;
    p2=0;
    ** calculated probabilities of second stage enrollment **;
    min=max(0, &di);
    max=min(&ni1, &n01+&di);
    if max<&ni1 then do;
        p1=p1+(1-probbnml(&p, &ni1,max));
        p2=p2+(1-probbnml(&pa,&ni1,max));
    end;
    do n=min to max;
        if n>0 then do;
            p1=p1+(probbnml(&p,&ni1,n)-probbnml(&p,&ni1,n-1))
                *probbnml(&p,&n01,n-&di);
            p2=p2+(probbnml(&pa,&ni1,n)-probbnml(&pa,&ni1,n-1))
                *probbnml(&p,&n01,n-&di);
        end;
        else if n=0 then do;
            p1=p1+probbnml(&p,&ni1,n)*probbnml(&p,&n01,n-&di);
            p2=p2+probbnml(&pa,&ni1,n)*probbnml(&p,&n01,n-&di);
        end;
    end;
    call symput("p1",trim(left(put(p1,8.7))));
    call symput("p2",trim(left(put(p2,8.7))));
    err2=1-probnorm(&&cr&dose);
    err3=1-probnorm(&&cr&pdose);
    p95_2=(1-err2/p1)*100;
    p95_3=(1-err3/p1)*100;

```



```

call symput("p1", trim(left(put(p1,8.7))));
call symput("p2", trim(left(put(p2,8.7))));
call symput("p95_2",trim(left(put(p95_2,8.7))));
call symput("p95_3",trim(left(put(p95_3,8.7))));
if p1>err2 and p1>err3 and p1<0.99 then call symput("va","1");
run;

%if %substr(&va,1,1)=1 %then %do;

** derive simulated sample of testing statistics **;
data tmp2;
  set rn1;
  if m+&di>0 then do;
    n=m+&di-.5;
    lambda=((n+0.5)/(&n1-n+0.5)) / ((m+.5)/(&n01-m+.5));
  end;
  else if m+&di<=0 then lambda=0.00000001;
  zi1=z1*&s1;
  if zi1>=log(lambda) then do;
    zi=&rho*&s2*z1+sqrt(1-&rho*&rho)*&s2*z2;
    output;
  end;
run;

proc univariate data=tmp2 noprint;
  var zi;
  output out=out pctlpts=&p95_2 &p95_3 pctlpre=pct_ pctlname=pt2 pt3;
run;

data _null_;
  set out;
  call symput("pct_pt2", trim(left(put(pct_pt2,10.6))));
  call symput("pct_pt3", trim(left(put(pct_pt3,10.6))));
run;

data _null_;
  set rn2 end=last;

```

```

retain err2 err3;
if _n_=1 then do;
    err2=0;
    err3=0;
end;
if mi1-m01>=&di then do;
    m0=m01+m02;
    mi=mi1+mi2;
    odds=((mi+0.5)/(&ni-mi+0.5))/((m0+0.5)/(&n0-m0+0.5));
    z=log(odds);
    if z>&pct_pt2 then err2=err2+1;
    if z>&pct_pt3 then err3=err3+1;
end;
if last then do;
    call symput("err2", trim(left(put(err2/&msz,10.6))));
    call symput("err3", trim(left(put(err3/&msz,10.6))));
end;
run;

data _null_;
    ** calculate power of the dose level with adaptive sampling **;
    set rn3 end=last;
    retain power2 power3;
    if _n_=1 then do;
        power2=0;
        power3=0;
    end;
    if mi1-m01>=&di then do;
        m0=m01+m02;
        mi=mi1+mi2;
        odds=((mi+0.5)/(&ni-mi+0.5))/((m0+0.5)/(&n0-m0+0.5));
        z=log(odds);
        if z>&pct_pt2 then power2=power2+1;
        if z>&pct_pt3 then power3=power3+1;
    end;
    if last then do;
        call symput("power2", trim(left(put(power2/&msz,10.6))));

```

```

    call symput("power3", trim(left(put(power3/&msz,10.6))));
end;
run;

/*****
g01: gain when current dose not effective and use adaptive
g02: gain when current dose not effective and only enroll first stage
g03: gain when current dose not effective and enroll all subjects
g11: gain when current dose effective and use adaptive
g12: gain when current dose effective and enroll all subjects
*****/

data tmp2;
  ** calculate maximum opportunity losses **;
  dose=&dose;
  p1=&p1;
  p2=&p2;
  di=&di;
  err2=&err2;
  err3=&err3;
  power2=&power2;
  power3=&power3;
  pct_pt2=&pct_pt2;
  pct_pt3=&pct_pt3;
  pow2=probnorm((&mu-&&cr&dose*&s2)/&s2);
  pow3=probnorm((&mu-&&cr&pdose*&s3)/&s3);
  err=probnorm(-&&cr&dose);
  %if &dose=2 %then %do;
    g01=err*p1+(1-err)*pow2*p1+pow3*(1-p1);
    g11=(power2*&r2+p2*(1-power2/p2)*pow2*&r3+(1-&p2)*pow3*&r3)/&r2;
    g02=pow3;
    g12=(pow2*&r2+pow2*(1-pow2)*&r3)/&r2;
    g03=err+(1-err)*pow2;
    l0=g02-g01;
    l1=g12-g11;
    g_min=min(g01, g11);
    l_max=max(l0, l1);
  %end;

```

```

%end;
%else %if &dose=1 %then %do;
  g01=p1*((1-err/p1)*(1-pow2)*pow2*&r3+(1-err/p1)*pow2*&r2+err/p1*&r2)
    /&r2+(1-p1)*&&g1_&pdose;
  g02=&&g1_&pdose;
  g11=p2*((1-power2/p2)*(1-pow2)*pow2*&r3+(1-power2/p2)*pow2*&r2
    +&power2/p2*&r1)/&r1+(1-p2)*&&g1_&pdose*&r2/&r1;
  g12=((1-pow2)**2*pow2*&r3+(1-pow2)*pow2*&r2+pow2*&r1)/&r1;
  g03=(err*&r2+(1-err)*pow2*&r2+(1-err)*(1-pow2)*pow2*&r3)/&r2;
  l0=g02-g01;
  l1=g12-g11;
  g_min=min(g01, g11);
  l_max=max(l0, l1);
%end;
run;

data report;
  set report tmp2;
run;

%end;
%end;

data report;
  set report;
  if _n_=1 then delete;
run;

proc sort data=report out=sorted;
  by l_max;
run;

proc print data=report noobs;
  var dose di p1 p2 g11 g01 g03 g12 l_max;
run;

** pick the design with minimum maximum opportunity loss **;

```

```

data _null_;
  set sorted(obs=1);
  call symput("pcut&dose", trim(left(put(p1, 6.5))));
  call symput("g0_&dose", trim(left(put(g01, 6.5))));
  call symput("g1_&dose", trim(left(put(g11, 6.5))));
run;

data design;
  set design sorted(obs=1);
run;

%end;

title "P=&p Pa=&pa Sample sizes=&ni1 &nni weights=&r3 &r2 &r1";

proc print data=design noobs;
  where dose>.;
  var dose di p1 p2 g11 g01 g03 l_max;
run;

%mend;

%dsn(p=0.1, pa=0.14, ni1=200, n01=200, nn0=1000, nni=1000, msz=200000,
  r3=1, r2=2, r1=3);
%dsn(p=0.1, pa=0.14, ni1=200, n01=200, nn0=1000, nni=1000, msz=200000,
  r3=1, r2=3, r1=5);
%dsn(p=0.1, pa=0.14, ni1=200, n01=200, nn0=1000, nni=1000, msz=200000,
  r3=1, r2=3, r1=7);
%dsn(p=0.1, pa=0.14, ni1=200, n01=200, nn0=1000, nni=1000, msz=200000,
  r3=1, r2=4, r1=9);

%dsn(p=0.1, pa=0.14, ni1=300, n01=300, nn0=1000, nni=1000, msz=200000,
  r3=1, r2=2, r1=3);
%dsn(p=0.1, pa=0.14, ni1=300, n01=300, nn0=1000, nni=1000, msz=200000,
  r3=1, r2=3, r1=5);
%dsn(p=0.1, pa=0.14, ni1=300, n01=300, nn0=1000, nni=1000, msz=200000,
  r3=1, r2=3, r1=7);

```

```
%dsn(p=0.1, pa=0.14, ni1=300, n01=300, nn0=1000, nni=1000, msz=200000,  
r3=1, r2=4, r1=9);
```

```
%dsn(p=0.1, pa=0.14, ni1=400, n01=400, nn0=1000, nni=1000, msz=200000,  
r3=1, r2=2, r1=3);
```

```
%dsn(p=0.1, pa=0.14, ni1=400, n01=400, nn0=1000, nni=1000, msz=200000,  
r3=1, r2=3, r1=5);
```

```
%dsn(p=0.1, pa=0.14, ni1=400, n01=400, nn0=1000, nni=1000, msz=200000,  
r3=1, r2=3, r1=7);
```

```
%dsn(p=0.1, pa=0.14, ni1=400, n01=400, nn0=1000, nni=1000, msz=200000,  
r3=1, r2=4, r1=9);
```

## APPENDIX D

### SAS PROGRAMS USED TO OBTAIN THE OPTIMAL CUT POINTS WHEN STEPDOWN TESTING APPROACHED IS USED

```
options symbolgen mprint mlogic;
%macro dsn(p=, pa=, ni1=, n01=, nni=, nn0=, msz=, r3=, r2=, r1=, seed=1009);
data design;
  dose=.; di=.; p1=.; p2=.; pct_med=.; err95=.; power95=.; g11=.; g01=.;
  g03=.; l1=.; g01=.; g02=.; l0=.; l_max=.; g_min=.;
run;
%do i=1 %to 2;
data report;
  dose=.; di=.; p1=.; p2=.; pct_med=.; err95=.; power95=.; g11=.; g01=.;
  g03=.; l1=.; g01=.; g02=.; l0=.; l_max=.; g_min=.;
run;
%let dose=%eval(3-&i);
%let pdose=%eval(3-&i+1);
%let ni=%sysevalf((&nni-&ni1)*(&dose-1)/(3-&dose+1)+&nni);
%let n0=%sysevalf((&nn0-&n01)*(&dose-1)/(3-&dose+1)+&nn0);
%let ni2=%sysevalf((&nni-&ni1)*(&dose)/(3-&dose)+&nni);
%let n02=%sysevalf((&nn0-&n01)*(&dose)/(3-&dose)+&nn0);
data rn1 rn2 rn3;
  do i=1 to &msz;
    m=ranbin(&seed,&n01,&p);
    z1=rannor(&seed);
    z2=rannor(&seed);
    output rn1;
```

```

end;
do i=1 to &msz;
    m02=ranbin(&seed,&n0.-&n01,&p);
    mi2=ranbin(&seed,&ni.-&ni1,&p);
    m01=ranbin(&seed,&n01,&p) ;
    mi1=ranbin(&seed,&ni1,&p);
    output rn2;
end;
do i=1 to &msz;
    m01=ranbin(&seed,&n01,&p);
    mi1=ranbin(&seed,&ni1,&pa);
    m02=ranbin(&seed,&n0.-&n01,&p);
    mi2=ranbin(&seed,&ni.-&ni1,&pa);
    output rn3;
end;
run;

data _null_;
    mu=log(&pa/(1-&pa)/(&p/(1-&p)));
    s1=sqrt(1/(&p*&n01)+1/(&n01-&p*&n01)+1/(&p*&ni1)+1/(&ni1-&p*&ni1));
    s2=sqrt(1/(&p*&n0)+1/(&n0-&p*&n0)+1/(&p*&ni)+1/(&ni-&p*&ni));
    s3=sqrt(1/(&p*&n02)+1/(&n02-&p*&n02)+1/(&p*&ni2)+1/(&ni2-&p*&ni2));
    rho=(s2*s2)/(s1*s2);
    call symput("s1", trim(left(put(s1,10.6))));
    call symput("s2", trim(left(put(s2,10.6))));
    call symput("s3", trim(left(put(s3,10.6))));
    call symput("rho", trim(left(put(rho,10.6))));
    call symput("mu", trim(left(put(mu,10.6))));
run;

%do di=-50 %to 50;
%let va=0;
data _null_;
    p1=0;
    p2=0;
    min=max(0, &di);

```



```

max=min(&ni1, &n01+&di);
if max<&ni1 then do;
    p1=p1+(1-probbnml(&p, &ni1,max));
    p2=p2+(1-probbnml(&pa,&ni1,max));
end;
do n=min to max;
    if n>0 then do;
        p1=p1+(probbnml(&p,&ni1,n)-probbnml(&p,&ni1,n-1))
            *probbnml(&p,&n01,n-&di);
        p2=p2+(probbnml(&pa,&ni1,n)-probbnml(&pa,&ni1,n-1))
            *probbnml(&p,&n01,n-&di);
    end;
    else if n=0 then do;
        p1=p1+probbnml(&p,&ni1,n)*probbnml(&p,&n01,n-&di);
        p2=p2+probbnml(&pa,&ni1,n)*probbnml(&p,&n01,n-&di);
    end;
end;
call symput("p1",trim(left(put(p1,8.7))));
call symput("p2",trim(left(put(p2,8.7))));
if p1>0.05 and p1<0.99 then do;
    call symput("va","1");
    call symput("p95",trim(left(put((1-0.05/p1)*100,8.7))));
end;
run;

%if %substr(&va,1,1)=1 %then %do;

data tmp2;
    set rn1;
    if m+&di>0 then do;
        n=m+&di-.5;
        lambda=((n+0.5)/(&ni1-n+0.5)) / ((m+.5)/(&n01-m+.5));
    end;
    else if m+&di<=0 then lambda=0.00000001;
    zi1=z1*&s1;
    if zi1>=log(lambda) then do;
        zi=&rho*&s2*z1+sqrt(1-&rho*&rho)*&s2*z2;
    end;
end;

```

```

        output;
    end;
run;

proc univariate data=tmp2 noprint;
    var zi;
    output out=out pctlpts=&p95 pctlpre=pct_ pctlname=med;
run;

data _null_;
    set out;
    call symput("pct_med", trim(left(put(pct_med,10.6))));
run;

data _null_;
    set rn2 end=last;
    retain err95;
    if _n_=1 then err95=0;
    if mi1-m01>=&di then do;
        m0=m01+m02;
        mi=mi1+mi2;
        odds=((mi+0.5)/(&ni-mi+0.5))/((m0+0.5)/(&n0-m0+0.5));
        z=log(odds);
        if z>&pct_med then err95=err95+1;
    end;
    if last then call symput("err95", trim(left(put(err95/&msz,10.6))));
run;

data _null_;
    set rn3 end=last;
    retain power95;
    if _n_=1 then power95=0;
    if mi1-m01>=&di then do;
        m0=m01+m02;
        mi=mi1+mi2;
        odds=((mi+0.5)/(&ni-mi+0.5))/((m0+0.5)/(&n0-m0+0.5));
        z=log(odds);

```

```

    if z>&pct_med then power95=power95+1;
end;
if last then call symput("power95", trim(left(put(power95/&msz,10.6))));
run;

```

```

/*****
g01: gain when current dose not effective and use adaptive
g02: gain when current dose not effective and only enroll first stage
g03: gain when current dose not effective and enroll all subjects
g11: gain when current dose effective and use adaptive
g12: gain when current dose effective and enroll all subjects
*****/

```

```

data tmp2;
  dose=&dose;
  p1=&p1;
  p2=&p2;
  di=&di;
  err95=&err95;
  pct_med=&pct_med;
  power95=&power95;
  pow2=probnorm((&mu-probit(0.95)*&s2)/&s2);
  pow3=probnorm((&mu-probit(0.95)*&s3)/&s3);
  %if &dose=2 %then %do;
    g01=pow2*&p1+pow3*(1-&p1);
    g11=(p2*pow2*(1-&power95/p2)*&r3+p2*pow2*(&power95/p2)*&r2+(1-&p2)
      *pow3*&r3)/&r2;
    g02=pow3;
    g12=(pow2*(1-pow2)*&r3+pow2**2*&r2)/&r2;
    g03=pow2;
    l0=g02-g01;
    l1=g12-g11;
    g_min=min(g01, g11);
    l_max=max(l0, l1);
  %end;
  %else %if &dose=1 %then %do;
    g01=p1*(pow2*(1-pow2)*&r3+pow2**2*&r2)/&r2+(1-p1)*&g1_&dose;
  %end;

```

```

g02=&&g1_&pdose;
g03=(pow2*(1-pow2)*&r3+pow2**2*&r2)/&r2;
g11=p2*(pow2*(1-pow2)*&r3+pow2**2*(1-pow2**2)*&r2+pow2**2
    *&power95/p2*&r1)/&r1+(1-p2)*&&g1_&pdose*&r2/&r1;
g12=(pow2*(1-pow2)*&r3+pow2**2*(1-pow2)*&r2+pow2**3*&r1)/&r1;
l0=g02-g01;
l1=g12-g11;
g_min=min(g01, g11);
l_max=max(l0, l1);
%end;
run;

data report;
  set report tmp2;
run;

%end;
%end;

data report;
  set report;
  if _n_=1 then delete;
run;

proc sort data=report out=sorted;
  by l_max;
run;

proc print data=report noobs;
  var dose di p1 p2 g11 g01 g12 g03 l_max;
run;

data _null_;
  set sorted(obs=1);
  call symput("pcut&dose", trim(left(put(p1, 6.5))));
  call symput("g0_&dose", trim(left(put(g01, 6.5))));
  call symput("g1_&dose", trim(left(put(g11, 6.5))));

```

```

    call symput("g2_&dose", trim(left(put(g03, 6.5))));
run;

title "P=&p Pa=&pa      sample sizes=&ni1 &nni      weights=&r3 &r2 &r1";
data design;
    set design sorted(obs=1);
run;

%end;

proc print data=design noobs;
    where dose>.;
    var dose di p1 p2 g11 g01 g03 l_max;
run;

%mend;

%dsn(p=0.1, pa=0.14, ni1=200, n01=200, nn0=1000, nni=1000, msz=400000,
    r3=1, r2=2, r1=3);
%dsn(p=0.1, pa=0.14, ni1=200, n01=200, nn0=1000, nni=1000, msz=400000,
    r3=1, r2=3, r1=5);
%dsn(p=0.1, pa=0.14, ni1=200, n01=200, nn0=1000, nni=1000, msz=400000,
    r3=1, r2=3, r1=7);
%dsn(p=0.1, pa=0.14, ni1=200, n01=200, nn0=1000, nni=1000, msz=400000,
    r3=1, r2=4, r1=9);

%dsn(p=0.1, pa=0.14, ni1=300, n01=300, nn0=1000, nni=1000, msz=400000,
    r3=1, r2=2, r1=3);
%dsn(p=0.1, pa=0.14, ni1=300, n01=300, nn0=1000, nni=1000, msz=400000,
    r3=1, r2=3, r1=5);
%dsn(p=0.1, pa=0.14, ni1=300, n01=300, nn0=1000, nni=1000, msz=400000,
    r3=1, r2=3, r1=7);
%dsn(p=0.1, pa=0.14, ni1=300, n01=300, nn0=1000, nni=1000, msz=400000,
    r3=1, r2=4, r1=9);

%dsn(p=0.1, pa=0.14, ni1=400, n01=400, nn0=1000, nni=1000, msz=400000,

```

```
    r3=1, r2=2, r1=3);  
%dsn(p=0.1, pa=0.14, ni1=400, n01=400, nn0=1000, nni=1000, msz=400000,  
    r3=1, r2=3, r1=5);  
%dsn(p=0.1, pa=0.14, ni1=400, n01=400, nn0=1000, nni=1000, msz=400000,  
    r3=1, r2=3, r1=7);  
%dsn(p=0.1, pa=0.14, ni1=400, n01=400, nn0=1000, nni=1000, msz=400000,  
    r3=1, r2=4, r1=9);
```

## BIBLIOGRAPHY

- [1] Peter Bauer, *A note on multiple testing procedures in dose finding*, *Biometrics* **53** (1997), 1125–1128.
- [2] Peter Bauer, Werner Brannath, and Martin Posch, *Multiple testing for identifying effective and safe treatments*, *Biometrical Journal* **5** (2001), 605–616.
- [3] Robert Bohrer, *On sharpening scheffe bounds*, *Journal of the Royal Statistical Society* **29** (1967), 110–114.
- [4] Frank Bretz, Ludwig A. Hothorn, and Jason C. Hsu, *Identifying effective and/or safe doses by stepwise confidence intervals for ratios*, *Statistics in Medicine* **22** (2003), 847–858.
- [5] Lawrence D. Brown, T. Tony Cai, and Anirban DasGupta, *Interval estimation for a binomial proportion*, *Statistical Science* **16** (2001), 101–133.
- [6] Edward J. Channon and Damian J. McEntegart, *Confidence intervals and p-values for williams' and other step-down multiple comparison tests against control*, *Journal of Biopharmaceutical Statistics* **11** (2001), 45–63.
- [7] T. Timothy Chen, *Optimal three-stage designs for phase ii cancer clinical trials*, *Statistics in Medicine* **16** (1997), 2701–2711.
- [8] Charles W. Dunnett, *A multiple comparison procedure for comparing several treatments with a control*, *Journal of the American Statistical Association* **50** (1955), 1096–1121.
- [9] ———, *New tables for multiple comparison with a control*, *Biometrics* **20** (1964), 482–491.
- [10] P. I. Feder, C. T. Olson, D. W. Hobson, M. C. Matthews, and R. L. Joiner, *Stagewise, group sequential experimental designs for quantal responses - one-sample and two-sample comparisons*, *Neuroscience and Biobehavioural Reviews* **15** (1991), 129–133.
- [11] John J. Hanfelt, *Optimal multi-stage designs for a phase ii trials that permits one dose escalation*, *Statistics in Medicine* **18** (1999), 1323–1339.
- [12] A. J. Hayter, *A one-sided studentized range test for testing against a simple ordered alternatives*, *Journal of the American Statistical Association* **85** (1990), 778–785.
- [13] Christopher Jennison and Bruce W. Turnbull, *Group sequential methods with applications to clinical trials*, Chapman & Hall / CRC, New York, 2000.

- [14] E. L. Korn, *Confidence bands for isotonic dose-response curves*, *Applies Statistics* **31** (1982), 59–63.
- [15] Chu-In Charles Lee, *On estimation for monotone dose-response curves*, *Journal of the American Statistical Association* **91** (1996), 1110–1119.
- [16] Denis Heng-Yan Leung and You-Gan Wang, *Isotonic designs for phase i trials*, *Controlled Clinical Trials* **22** (2001), 126–138.
- [17] Ruth Marcus, *Some results on simultaneous confidence intervals for monotone contrasts in one-way anova model*, *Communications in Statistics - Theory and Method* **11** (1982), 615–622.
- [18] Ruth Marcus, Eric Peritz, and K. R. Gabriel, *On closed testing procedures with special reference to ordered analysis of variance*, *Biometrics* **63** (1976), 655–660.
- [19] Ruth Marcus and Eric Pertiz, *Some simultaneous confidence bounds in normal models with restriced alternatives*, *Journal of the Royal Statistical Society* **38** (1976), 157–165.
- [20] Allan Pally, *A decisian analytic approach to a futility analysis of a phase ii pharmaceutical study*, *Journal of Biopharmaceutial Statistics* **11** (2001), 209–225.
- [21] Daniel J. Schaid, Sam Wieand, and Terry M. Therneau, *Optimal two-stage screening designs for survival comparisons*, *Biometrika* **77** (1990), 507–513.
- [22] David A. Schoenfeld, *Confidence bounds for normal means under order restrictions, with application to dose-response curves, toxicology experiments, and low-dose extrapolation*, *Journal of the American Statistical Association* **81** (1986), 186–195.
- [23] Richard Simon, *Optimal two-stage designs for phase ii clinical trials*, *Controlled Clinical Trials* **10** (1989), 1–10.
- [24] Paul N. Somerville, *Multiple testing and simultaneous confidence intervals: calculation of constants*, *Computational statistics & Data analysis* **25** (1997), 217–233.
- [25] ———, *Obtaining critical values for simultaneous confidence intervals and multiple testing*, *Biometrical Journal* **43** (2001), 657–663.
- [26] Nigel Stallard, *Decision-theoretic designs for phase ii clinical trials allowing for competing studies*, *Biometrics* **59** (2003), 402–409.
- [27] ———, *Decision theoretic designs for phase ii clinical trials allowing for competing studies*, *Biometrics* **59** (2003), 402–409.
- [28] Ajit C. Tamhane, Yosef Hochberg, and Charles W. Dunnett, *Multiple testing procedures for dose finding*, *Biometrics* **52** (1996), 21–37.
- [29] Ajit C. Tamhane and Brent R. Logan, *Multiple testing procedures for indentifying the minimum effective and maximum safe doses of a drug*, *Journal of the American Statistical Association* **97** (2002), 293–301.
- [30] Martin A. Tanner, *Tools for statistical inference*, Springer-Verlag, New York, 1996.



- [31] Peter F. Thall, Richard Simon, and Susan S. Ellenberg, *Two-stage selection and testing designs for comparative clinical trials*, *Biometrika* **75** (1988), 303–310.
- [32] ———, *A two-stage design for choosing among several experimental treatments and a control in clinical trials*, *Biometrics* **45** (1989), 537–547.
- [33] John Whitehead, *The design and analysis of sequential clinical trials*, John Wiley & Sons, Inc., Chichester, 1997.
- [34] D. A. Williams, *A test for differences between treatment means when several dose levels are compared with a zero dose level*, *Biometrics* **27** (1971), 103–117.
- [35] ———, *The comparison of several dose levels with a zero dose control*, *Biometrics* **28** (1972), 519–531.
- [36] ———, *Some inference procedures for monotonically ordered normal means*, *Biometrika* **64** (1977), 9–14.