DOSE FINDING STRATEGIES FOR SINGLE DRUG AND COMBINATION DRUG TRIALS

by

Julia Soulakova

M.A., Wayne State University, 2001

B.S., Moscow State University, 1998

Submitted to the Graduate Faculty of

the Faculty of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2006

UNIVERSITY OF PITTSBURGH

FACULTY OF ARTS AND SCIENCES

This dissertation was presented

by

Julia Soulakova

It was defended on

July18, 2006

and approved by

Henry W. Block, Professor, Department of Statistics

Leon J. Gleser, Professor, Department of Statistics

Allan R. Sampson, Professor, Department of Statistics

Lisa Weissfeld, Professor, Department of Biostatistics

Dissertation Advisor: Allan R. Sampson, Professor, Department of Statistics

DOSE FINDING STRATEGIES FOR SINGLE DRUG AND COMBINATION DRUG TRIALS

Julia Soulakova, PhD

University of Pittsburgh, 2006

A key component of drug development is to establish the compound's dose-response relationship, and identify all effective doses of the drug with a general goal of selecting the minimum effective dose (MED).

A new closed testing procedure is proposed for identifying the MED for a single component drug. This procedure is based on constructing simultaneous one-sided confidence bands for the response surface of each dose's effect relative to placebo. Our methodology utilizes a stepwise closed testing to test the ordered hypotheses of equality of mean dose-responses. The pattern of the rejected and accepted null hypotheses provides the estimate of the MED, if it exists.

In the case of a combination drug, in addition to demonstrating safety and efficacy the FDA requires demonstrating that each component makes a contribution to the claimed effects. A combination which satisfies the last requirement is called an efficacious combination.

In the most common case both single drugs are approved ones, and therefore, the efficacious combinations are effective, that is, they produce a therapeutic effect which is superior to placebo.

We propose a closed testing procedure for estimating the minimum efficacious combinations (MeD's) in a two-drug study and introduce a notion of the MeD-set.

The main advantage of a closed testing procedure is the strong control of the familywise error at level of significance α and allowing testing individual hypotheses at the same significance level α without multiplicity adjustments.

The proposed procedure is based on two main steps. In the first step, all possible structures of the population MeD-set are identified and the related closed family of hypotheses is constructed, and the proper step-down testing partial order is established. The second step is the

"α-testing" step. Using the closed testing principle, we test the hypotheses by constructing the AVE-test statistic. The pattern of the rejected null hypotheses identifies the MeD-set.

In order to assess the performance of our procedure, we define several statistical measures. These notions are used in a large simulation study to examine the goodness of the estimation procedures and to identify the population configurations when the procedure performs the best.

Keywords: Minimum effective dose; Dose-response; Combination drug; AVE-test; Closed testing; Step-down procedure.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

My father says that "there are just a few important decisions that a person must make during his life but upon these decisions his whole life depends." Choosing a field of the major interest and a school where one can get an excellent education in this field have a great effect on the person's life-time career and are two of these choices.

My first choice to become a mathematician was made while I was in the middle school by my whole family. Since my mother and my sister are chemists and my father is an engineer, this decision was purely my father's idea. It is never easy to argue with a woman, and it is almost impossible to argue with two of them at once, so *the battle* was a hard one but my father won it! More likely, he provided a very complex proof that his choice is the only correct one. Moreover, I suspect that the proof was so *numerically intense* that no one wanted to spend time to verify it, so the decision was made. This is how *I chose* the field of my interest!

Next, my sister chose the only university for me to go to, Moscow State University. And again, my whole family helped me a lot so I could pass the exams and get admitted at such a famous school. This is how *I made* the second important decision in my life!

I am really thankful to the faculty members of the Department of Computational Mathematics and Cybernetics of Moscow State University for their unique teaching style and creating an atmosphere stimulating a self-education and individual growth as a scientist of each student.

My second school, Wayne State University, was also chosen for me by my sister. While I was an undergraduate student, my sister decided to continue her graduate education in the United States. No one in my family had ever thought about leaving Russia but my sister who is very brave and enthusiastic about favoring new opportunities. Following her steps, I started working

on my Masters' at Wayne State University. This is how *I made* the third important decision in my life!

I am very thankful to all faculty members of the Department of Mathematics at the Wayne State University. Especially, I want to thank two Professors, Rafail Khasminskii and Alex Korostelev who helped me make a decision to continue my education in the field of statistics.

My third school, the University of Pittsburgh was not chosen for me by my sister. This time I had to make my own choice and I made a very right one. During my second year as a graduate student at the Department of Statistics I asked Professor Allan R. Sampson to become my research advisor. I was and I am very lucky to have his support and time for more than four years. And the most importantly, I am very thankful to Allan R. Sampson for helping me in selecting a dissertation topic, guiding me during my research and playing the leading role in my early career achievements as a young statistician.

I also would like to acknowledge the help of other faculty members of the Department of Statistics, particularly, Professors Leon J. Gleser and Henry W. Block for being the dissertation committee members and giving me a professional advice, and the Chair of the department, Professor Satish Iyengar, for creating an educational atmosphere at the department. I also would like to thank two Professors of the Department of Biostatistics, John Bryant and Lisa Weissfeld, for their valuable comments regarding my research.

Finally, I would like to thank my very closest family members: my husband, Alex Goponenko, for his love, patience, support, and motivation, and two of my loving children, Ksenia and Ivan, for filling every moment of my life with joy and happiness.

"There are just a few important decisions that a person must make during his life;" and maybe, the majority of these decisions were made for me, I do not regret any of them. I believe that without my family and my teachers I would not accomplish that much and I would not become a proud owner of a *Philosophiæ Doctor* degree.

1.0 INTRODUCTION

Modern drug development focuses on assessing the biological activity of a chemical or biological compound and investigating the significant health benefits and adverse effects, due to the compound. The Food and Drug Administration (FDA) requires demonstrating the safety and efficacy of the drug for approval for marketing. A key component of drug development is to establish the compound's dose-response relationship, and based on this relationship identify all effective doses of the drug with a general goal of selecting the lowest dose among the effective ones. This information is of significance, particularly as increasing dosage tends to lead to the increased serious side effects. The recent actions concerning Vioxx® (Merck) or Celebrex® (Pfizer) illustrate the importance of understanding the dose-response relationship, particularly with respect to safety. The problems in establishing the correct dose-response relationships are even more complex when combination drugs are considered, such as the cholesterol drug Vytorin® (Merck/Schering-Plough), which is a combination of Zocor and Zetia®.

The first results presented in this dissertation deal with single drug studies. The statistical analyses of single drug dose-response studies, while well developed, are still challenging. In a typical clinical trial, subjects are randomly assigned to one of the several dose groups and administered a drug at the assigned dose level. The patients in the control group receive placebo. All patients are treated and observed over the same amount of time. The main goal of the trial is to draw statistical inferences concerning the behavior of the population dose response function. Of particular importance is to determine the dose range where the drug is effective and then to obtain the lowest dose that is still effective. If the true dose-response were known then all of those notions would be simple. Data from clinical trials are used to estimate the shape of the dose-response. After the dose-response relationship is estimated, another goal is to estimate all effective doses, with a primary emphasis on estimation of the minimum effective dose (MED) and to characterize the uncertainty of this estimate.

There are a number of well known statistical procedures that are used to estimate the MED, but there is no one best procedure since the properties of the procedure depend on the shape of the true, but unknown, dose-response curve. A literature review (given in Section 2.1) discusses various hypothesis testing procedures and methods involving suitable confidence bands for the dose-response curve. We present a new procedure for identifying the MED in single drug studies. In doing so, we develop a new procedure for constructing simultaneous one-sided and two-sided confidence bands for the dose-response curve of each dose's effect relative to placebo. Our methodology utilizes a method similar to the approach given by Lee (1996). By generalizing Lee's method to handle dependent random variables, we first construct two-sided confidence bands for the differences between responses to the active dose levels and placebo (see Section 2.3). Next, we construct the lower bands for the differences between responses to the active dose levels and placebo. Then we use stepwise closed testing to test the ordered hypotheses of equality of mean dose-responses. The pattern of the rejected and accepted null hypotheses provides the estimate of the MED, if it exists. The detailed step-down procedure is given in Section 2.4. We perform simulation studies to assess the goodness of the parameter estimates provided by the procedure. We also compare the proposed procedure with two well known procedures: William's (1971, 1972) procedure and the Step-Down with Pairwise Contrasts procedure considered by Tamhane *et al.* (1996) in order to identify in comparison dose-response patterns where our procedure performs well in identifying the MED and has a smaller bias. The results of these studies are presented in Section 2.5.

A much more complicated issue concerning dose-response studies deals with compounds obtained as a combination of two known effective drugs. There are many reasons why it may be desirable to produce a drug combination. There is a particular advantage to consider a drug combination for the treatment of infectious diseases. Wertheiner and Morrison (2002) note that simplifying the treatment regiment is crucial in the case of infectious diseases, because "partial adherence can lead to the development of drug-resistant strains and a threat to public health."

In the case of a combination drug, the FDA requires demonstrating that each component makes a contribution to the claimed effects. A combination which satisfies this requirement is called efficacious. The lowest efficacious combinations are called the minimum efficacious combinations (MeD's). In the most commonly occurring case, both single drugs are approved

ones. In such a case the MeD is effective, that is, it produces a therapeutic effect which is superior to placebo.

Since the MeD is not usually unique in a combination drug study, we define the notion of the MeD-set. It turns out that the population MeD-set has a very complicated structure and has a number of interesting properties, which are discussed in Section 3.2.

Due to the increased attention directed toward combining two or more therapeutic agents to produce better treatments, several procedures have recently been proposed by Hung (1992, 1993) for detecting the superior combinations and by Hellmich and Lehmacher (2005) for detecting the minimum effective combinations, *i.e.*, the combinations with higher mean dose-responses than the placebo. These procedures are discussed in a literature review provided in Section 4.1. But none of these methods was extended to find the MeD's.

As in a single drug study, in a two-drug study subjects are randomly allocated to the different groups, each group receiving a particular combination of the drugs. The increasing dose levels of both drugs are specified prior to the experiment. The patients in the control groups get either placebo or a particular dose of the component drug alone. All patients are treated and observed over the same amount of time. Because we are assuming in our research that each component drug is effective we do not use the data of the placebo group other than for the variance estimate.

We propose a step-down procedure for estimating the minimum efficacious doses (MeD's) in such a two-drug study. For each active compound, the expected gain from combining the drugs at the given levels is defined. Next, the appropriate estimator of the expected gain is constructed in order to test the hypotheses of the expected gains being zero versus at least one positive gain. The pattern of the rejected null hypotheses identifies all the combinations, *i.e.*, the MeD's, such that decreasing at least one component's dose leads to a no-gain compound. Our approach focuses on developing criteria necessary for proper ordering of the hypotheses of homogeneity of mean dose-responses and applying a suitable closed testing procedure. A general method for constructing a closed hypothesis family and ordering the hypotheses is discussed in Section 4.2. Since the set of hypotheses to be tested depends on the number of dose levels of each drug considered in the study, we focus on the methods for the 2 by 2 case, 2 by 3 case, and 3 by 3 case, which are presented, respectively, in Sections 4.3-4.5.

In order to assess the performance of the proposed procedure and any other procedure which identifies the MeD-set, several measures are considered. Some of them are generalizations of the measures used in a single drug case, such as the Familywise Error, Power and Lack of Power. In addition we define a number of other measures that can be used to assess the goodness of the parameter estimates. Chapter 5.0 is devoted to the criteria for evaluating procedures and presents the definitions of these measures.

We perform simulation studies to examine the performance of the proposed procedure. The design and results of the studies are presented in Chapter 6.0. For different population mean dose-response patterns for the 2 by 2 and 2 by 3 cases, the strong control of familywise error is confirmed. Also for the 2 by 3 case we compare two proposed step-down procedures, the regular and modified, and obtain the criteria when one procedure dominates the other in terms of the measures proposed in Chapter 5.0.

In summary, the contents of this dissertation are as follows. Chapter 2.0 is devoted to the single drug case. The literature review and introduction are presented in Section 2.1. Then the research problem is stated in Section 2.2. Section 2.3 presents the method for constructing simultaneous two-sided max-min confidence bands. A similar approach used to obtain one-sided bands is briefly discussed and then is utilized in the proposed procedure presented in Section 2.4. Section 2.5 concludes the single drug case by presenting the results of the simulation studies with regard to the proposed procedure.

The combination drug case is presented in Chapters 3.0-6.0. First, the design of combination drug studies and the related statistical parameters are discussed in Chapter 3.0. Then the proposed testing procedure is presented in Chapter 4.0 starting with the literature review given in Section 4.1. Chapter 5.0 discusses the criteria for evaluating the performance of a testing procedure applied in a combination drug setting. Chapter 6.0 presents the details of the design and results of the simulation studies. Finally, Chapter 7.0 discusses the findings presented in this dissertation and provides the main goals of the future research.

2.0 SINGLE DRUG STUDY

2.1 INTRODUCTION AND LITERATURE REVIEW

There are a number of different dose-response curves considered for modeling and analyses in a single drug study. Among the potential shapes, the monotone one is by far the most commonly discussed in the literature. Figure 1 illustrates two examples of monotone population dose-response functions: linear and step functions.

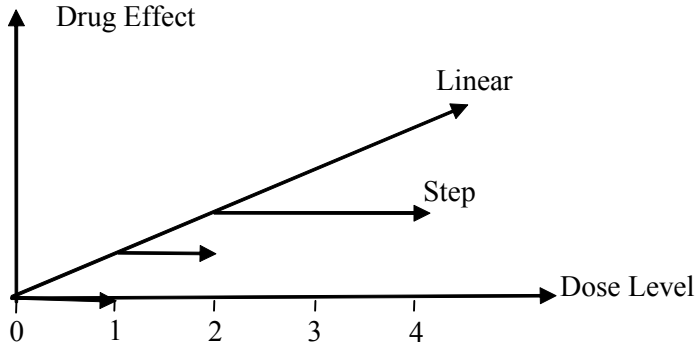

Figure 1. Examples of monotone dose-response curves

One of the most important statistical summaries of a single drug trial is the estimated dose-response curve along with the upper and lower confidence bands for the population curve. There are a number of procedures for estimation of the dose-response curve and obtaining those bounds. Lee (1996) proposed a procedure which uses the assumption of the monotone dose-response function and does not rely on a specified parametric model for the population curves. Before we discuss this procedure we introduce the basic notation and assumptions which we use.

In a single drug trial, subjects are randomly allocated to one of $K+1$ groups. Let n_i, $i = 0,\ 1,..., K$ denote the number of patients being allocated to the $i-th$ group. Subjects in each group receive a particular dose x_i of the drug, and the increasing dose levels $i = 0,\ 1,..., K$ are specified prior to the experiment. Patients in the control group receive placebo, denoted by dose $x_0 = 0$. Let $\mu_i = f(x_i)$ denote the population mean dose-response for the $i-th$ group. Then the assumption of a monotone non-decreasing dose-response function $f(x)$ can be written as

$$f(x_0) \le f(x_1) \le ... \le f(x_K) \tag{2.1}$$

Let y_{ij} denote the response of the $j-th$ subject in the $i-th$ group, $i = 0,\ 1,..., K$ and $j = 1,\ 2,...,\ n_i$. We assume that the following hold.

1) All observations y_{ij} are mutually independent. (2.2)

2) $y_{ij} \sim N(\mu_i, \sigma^2)$, with $\mu_i = f(x_i)$ for $i,\ i = 0,\ 1,..., K$ and $j = 1,\ 2,...,\ n_i$. (2.3)

Let $\bar{y}_i \sim N(\mu_i, \sigma^2 / n_i)$, $i = 0,\ 1,..., K$ denote the sample means of the observations at each dose level $i,\ i = 0,\ 1,..., K$. Let $\hat{\sigma}^2$ be an unbiased estimate of the variance σ^2 such that $\hat{\sigma}^2$ is independent of $\bar{y}_i$, $i = 0,\ 1,..., K$ and

$$\hat{\sigma}^2 \sim \sigma^2 \chi_\nu^2 / \nu, \text{ with } \nu = \sum_{i=0}^{K} n_i - (K+1) \text{ degrees of freedom.} \tag{2.4}$$

Sometimes it is convenient to assume that the same numbers of subjects are allocated to each treatment group, *i.e.*,

$$n_1 = n_2 = ... = n_K = n. \tag{2.5}$$

Under the conditions (2.1)-(2.4) Lee's (1996) simultaneous max-min confidence bands are given by

$$\max_{p \le q \le i} \left[\left(\sum_{j=p}^{q} \sqrt{n_j}\, \bar{y}_j - |t|_{K,\nu}^{\alpha} \hat{\sigma} \sqrt{q - p + 1} \right) \div \sum_{j=p}^{q} \sqrt{n_j} \right] \le f(x_i)$$

$$\le \min_{i \le r \le s} \left[\left(\sum_{j=r}^{s} \sqrt{n_j}\, \bar{y}_j + |t|_{K,\nu}^{\alpha} \hat{\sigma} \sqrt{s - r + 1} \right) \div \sum_{j=r}^{s} \sqrt{n_j} \right]$$

where $i = 0,\ 1,..., K$ and $|t|_{K,\nu}^{\alpha}$ denotes the $1-\alpha$ percentile of the distribution function of generalized studentized maximum modulus statistic T, defined as

$$T = \max_{0 \le i \le j \le K} \frac{\left| \sum_{h=i}^{j} z_h \right|}{\sqrt{j-i+1}} \Big/ \frac{\hat{\sigma}}{\sigma}, \tag{2.6}$$

where $z_h \sim N(0,1)$ for $h = 0,\ 1,..., K$ and are independent random variables.

We use ideas based on Lee's (1996) approach to construct max-min simultaneous confidence bounds for the differences between responses to the active dose levels and placebo. These two-sided bands can be used to assess the behavior of the dose response differences of a certain drug.

Jia (2004) introduces a method for obtaining the simultaneous confidence bands for the monotone mean differences $\nu_i = \mu_i - \mu_0 = f(x_i) - f(x_0),\ i = 1,\ 2,..., K$ for independent $\bar{y}_i \sim N(\mu_i, \sigma_i^2 / n_i),\ i = 0,\ 1,..., K$. To construct such intervals, Jia (2004) applies a max-min procedure introduced by Korn (1982) and constructs a step-down testing procedure for analyzing dose study data, which incorporates the adaptive sampling nature of the data.

Sampson *et al.* (2005) further develop the ideas of Lee (1996) and Korn (1982) by providing a unified presentation of Lee's and Korn's univariate simultaneous inference procedures for monotone functions. The graphical representation of simultaneous confidence bands allows for the exploration of the geometry of the procedures in the regression setting and, therefore, illustrates their feasibility and applicability. These authors also suggest several cases when the procedures can be extended and propose a new bandwidth procedure which generalizes Lee's and Korn's methods.

Next, let us consider two procedures for identifying the minimum effective dose (MED) presented in the literature, one of them assumes the monotone dose-response and the other does not. Both procedures use the following definitions of effective doses and MED.

The active drug dose x_i is considered effective if it has a mean dose-response higher than the placebo mean dose-response, *i.e.*, $f(x_0) < f(x_i)$, $1 \le i \le K$. If there is at least one effective dose among doses considered in the trial, then there is a minimum effective dose (MED), which is defined as follows. The dose x_m, $1 \le m \le K$ is the MED if $f(x_m) - f(x_0) > 0$ and $f(x_i) - f(x_0) \le 0$ *for* $i < m$, *i.e.*, $MED = \min_{1 \le i \le K} \{ i : \mu_i > \mu_0 \}$. One approach to the problem of identifying the MED is to formulate it as a sequence of hypothesis testing problems:

$$H_{0i} : \mu_0 = \mu_1 = ... = \mu_i , \tag{2.7}$$

versus the alternative $H_{ai} : \mu_0 = \mu_1 = ... = \mu_{i-1} < \mu_i$, $1 \le i \le K$. In this approach, if $i*$ is the smallest i such that H_{0i} is rejected, then $i*-th$ dose is declared to be the MED.

The main advantage of formulating the problem as a sequence of the above hypotheses is that the hypothesis family $\mathrm{H} = \{H_{0i}, i = 1,2,...K\}$ is a family that is closed under intersection, *i.e.*, any hypothesis which is obtained as an intersection of any number of the hypotheses belonging to the family belongs to H. Indeed, as shown by Tamhane *et al.* (1996), for any set of indexes $1 \le i_1 \le i_2 \le ... \le i_m \le K$, and any $H_{0i_s} \in \mathrm{H}$, $s = 1,2,...,m$ it follows that $H_{0i_1} \bigcap H_{0i_2} \bigcap ... \bigcap H_{oi_m} = H_{0i_m} \in \mathrm{H}$, so that the family is closed under intersection.

Marcus, Peritz and Gabriel (1976) propose a general method (hereafter denoted as the MPG principle) for testing hypotheses from a closed family. This method permits testing each hypothesis at level of significance α and states that a hypothesis is rejected at level α if and only if all hypotheses implying it, including itself, are significant at level α. This method has the nice property of strong control of familywise error (FWE) at level α. The FWE is defined as the probability of rejecting at least one true null hypothesis H_{0i}. Next, we review two procedures that use the MPG principle and, therefore, strongly control the FWE.

The problem of identifying the MED formulated by (2.7) can be handled by using Williams' (1971) procedure. This is a step-down procedure for dose-response functions satisfying (2.1). Under assumptions (2.2)-(2.4), the isotonic (maximum likelihood) estimates for the population mean responses are given by

$$\hat{\mu}_i = \max_{1 \le u \le i} \min_{1 \le v \le K} \frac{\sum_{i=u}^{v} \bar{y}_i}{v - u + 1}, \; 1 \le i \le K .$$

Next, under assumption of (2.5), pairwise t-type statistics are calculated as

$$t_i^W = \frac{\hat{\mu}_i - \bar{y}_0}{\hat{\sigma}\sqrt{1/n_0 + 1/n}}, \; 1 \le i \le K . \tag{2.8}$$

Hypotheses (2.7) are tested in a step-down manner by comparing the test statistics t_i^W's with the corresponding critical values. The critical values depend on the dose level i, degrees of freedom ν and level of significance α, and are given by Williams (1971, 1972). Starting from

$i = K$, H_{0K} is tested against H_{aK} at a level of significance α, if t_K^W is greater than or equal to the corresponding critical value, then H_{0K} is rejected and H_{0K-1} is tested. If H_{0K} is accepted then the testing is completed and all null hypotheses are accepted, otherwise H_{0K-1} is tested at level α. Similarly, if t_{K-1}^W is greater than or equal to its corresponding critical value, then H_{0K-1} is also rejected and so on; unless H_{0m} is accepted for some $m = 1,\ 2,..., K$, then all $H_{01}, H_{02},..., H_{0m}$ are accepted and MED is estimated as $MED^* = m + 1$.

Tamhane *et al.* (1996) mention the above procedure and also consider a class of stepwise closed testing procedures based on contrasts among the sample means. None of these latter procedures require the monotonicity assumption (2.1) and all of them strongly control the FWE. Here, we illustrate the step-down procedure under the assumptions of (2.2)-(2.5) which is based on the pairwise contrasts

$$t_i = \frac{\bar{y}_i - \bar{y}_0}{\hat{\sigma}\sqrt{1/n_0 + 1/n}},\quad i = 1,\ 2,..., K. \tag{2.9}$$

The joint distribution of these statistics is the multivariate $t-$distribution with ν degrees of freedom and correlation matrix $\{\rho_{ij}\}$, where $\rho_{ij} = n/(n_0 + n)$. In order to use the stepwise procedures, first, the statistics are ordered as $t_{(1)} \le t_{(2)} \le ... \le t_{(K)}$. Next, the hypotheses are also reordered as $H_{0(1)}, H_{0(2)},..., H_{0(K)}$, so that $H_{0(i)}$ corresponds to the statistic $t_{(i)}$, $1 \le i \le K$. The critical values $c_{i,\nu,\rho}^{\alpha}$ for testing $H_{0(i)}$ are obtained as the upper α equicoordinate critical points of the $i-$variate $t-$distribution with ν degrees of freedom and common correlation $\rho = \rho_{ij}$. The step-down procedure based on the pairwise contrasts (SDPC) is then applied in the same manner as the Williams' procedure but the order in which hypotheses are to be tested at the 1st step is specified by the $H_{0(K)}, H_{0(K-1)},..., H_{0(1)}$ instead of $H_{0K}, H_{0K-1},..., H_{01}$. If $t_{(K)} \ge c_{K,\nu,\rho}^{\alpha}$, then reject $H_{0(K)}$ and all hypotheses whose rejection is implied by it, e.g., if $H_{0(K)} = H_{0m}$ then reject all $H_{0j},\ \ j = m, m+1,..., K$ and proceed to the 2nd step to test $H_{01}, H_{02},..., H_{0m-1}$. If $t_{(K)} < c_{K,\nu,\rho}^{\alpha}$ then accept $H_{0(K)}$ and complete the testing. In general, in the $i-th$ step let k_i be the number of hypotheses still to be tested. Reorder the statistics $t_{(1)} \le t_{(2)} \le ... \le t_{(k_i)}$ and the corresponding hypotheses as $H_{0(1)}, H_{0(2)},..., H_{0(k_i)}$. Test $H_{0(k_i)}$ by comparing $t_{(k_i)}$ with $c_{k_i,\nu,\rho}^{\alpha}$. If $t_{k_i} \ge c_{k_i,\nu,\rho}^{\alpha}$,

then reject $H_{0(k_i)}$ and all hypotheses whose rejection is implied by it and go to the next step, otherwise, stop testing. When the testing procedure stops, the MED is estimated as the minimum index of the rejected hypotheses.

2.2 PROBLEM SET-UP

We are going to consider the problem of identifying the lowest dose level for which the mean response differs from the mean response at the placebo (zero) dose under the assumption of a monotone dose-response curve (2.1). At first, in addition to (2.2)-(2.4), we assume that σ^2 is known and that both of the following conditions hold

$$1)\ n_0 = n_1 = ... = n_K = n, \tag{2.10a}$$

$$2)\ 2\sigma^2 / n = 1. \tag{2.10b}$$

Let us construct the mean differences $\nu_i = \mu_i - \mu_0 = f(x_i) - f(x_0),\ i = 1,2,...,K$ with $\nu_0 = 0$.

Note that ν_i's satisfy the monotonicity assumption

$$0 \le \nu_1 \le \nu_2 \le ... \le \nu_K. \tag{2.11}$$

Then the minimum effective dose (MED) is defined as

$$MED = \min\ \{i : \mu_i > \mu_0\} = \min\ \{i : \nu_i > 0\}. \tag{2.12}$$

And the problem of identifying the MED is formulated as a sequence of hypothesis testing problems

$$H_{0l} : 0 = \nu_1 = \nu_2 = ... = \nu_l, \text{ versus the alternative}$$

$$H_{al} : 0 = \nu_1 = \nu_2 = ... = \nu_{l-1} < \nu_l,\ l = 1,2,...,K. \tag{2.13}$$

Obviously, the above family of hypotheses $H = \{H_{0l},\ l = 1,2,...,K\}$ is closed under intersection; the proof is identical to the one for hypothesis family (2.7) and is given in the preceding section. The last fact allows us to apply the MPG principle, and test each hypothesis at a level of significance α. As we mentioned above, such a testing procedure guarantees the strong control of the FWE. The detailed proposed procedure for identifying the MED is

described later on in Section 2.4.2. The next section illustrates our generalization of Lee's method to handle dependent random variables to construct two-sided confidence bands for the differences between responses to the active dose levels and placebo.

2.3 TWO-SIDED CONFIDENCE BANDS FOR MEAN DOSE-RESPONSE DIFFERENCES

2.3.1 Test Statistic and its Distributional Properties

Consider a vector of K random variables $(z_1, z_2, ..., z_K)'$ with z_h distributed marginally as standard normal, $h = 1,\ 2, ..., K$, and covariance $Cov(z_i, z_j) = \frac{1}{2}$ for $i \neq j$, *i.e.*, $(z_1, z_2, ..., z_K)' \sim MVN(\vec{0}_K, \Sigma)$, where $\vec{0}_K$ is the K-dimensional zero-vector and Σ is the $K \times K$ matrix

$$\Sigma = \begin{pmatrix} 1 & 0.5 & 0.5 & .... & 0.5 \\ 0.5 & 1 & 0.5 & .... & 0.5 \\ ..... & & & & \\ 0.5 & 0.5 & 0.5 & .... & 1 \end{pmatrix}.$$

Next, we compute the test statistics

$$T = \max_{1 \le i \le j \le K} \frac{\left| \sum_{h=i}^{j} z_h \right|}{\sqrt{j - i + 1}}. \qquad (2.14)$$

The distribution of this statistic is different from the distribution of the generalized standardized maximum modulus (2.6) because the z_h's are no longer independent. Note that T takes only positive values, and its distribution depends on K, the total number of z_h's. For $K = 1$, the statistic T reduces to the absolute value of the standard normal random variable and its distribution function can be expressed as

$$F_T(a) = \begin{cases} 2 \cdot \Phi(a) - 1, & a \geq 0 \\ 0, & a < 0 \end{cases}$$

where $\Phi(a)$ is the distribution function of a standard normal random variable evaluated at a, $a \in (-\infty, +\infty)$.

For $K >> 1$, the statistic does not have an easy expression for the density or the distribution function. The graph in Figure 2 illustrates the approximate density function obtained by simulation (based on 100,000 replications) for $K = 10$ under the assumptions (2.2)-(2.4) and (2.10).

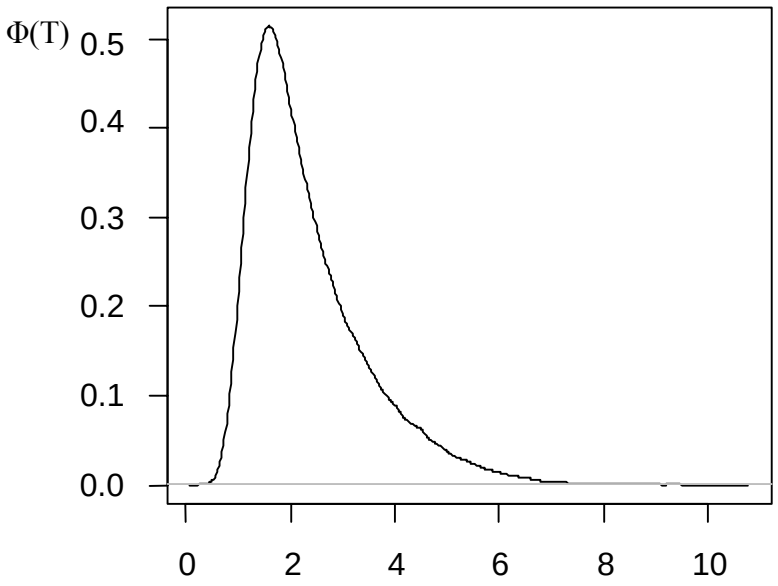

Figure 2. Approximated density function of T statistic (2.14) for K = 10

2.3.2 Constructing Two-Sided Simultaneous Max-Min Confidence Bands for the Case of Equal Sample Sizes and Known Variance

To construct the confidence bands, first we need to determine the percentage points of the distribution of T statistic. Let $|c|_K^{\alpha}$ denote the upper $1-\alpha$ percentile of the distribution of T for

K active dose levels. Since the distribution function does not have a simple expression, we approximate the percentiles by solving the equation below for $|c|_K^{\alpha}$

$$P(T \le |c|_K^{\alpha}) = 1 - \alpha \,. \tag{2.15}$$

The R code for obtaining the critical values based on a simulation study is given in Appendix A. The results for up to ten dose levels for $\alpha = 0.1,\ \alpha = 0.05$ and $\alpha = 0.01$ are given in Table 1. The number of replications for each case is 100,000.

Table 1. α -level critical values of T statistic given by (2.14)

Number of Active	Significance Level		
Dose Levels	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$
K=1	1.64	1.96	2.58
K=2	2.12	2.47	3.19
K=3	2.45	2.85	3.67
K=4	2.73	3.18	4.09
K=5	2.97	3.47	4.49
K=6	3.20	3.73	4.86
K=7	3.39	3.99	5.18
K=8	3.59	4.21	5.49
K=9	3.78	4.45	5.77
K=10	3.94	4.65	6.08

Under conditions (2.2)-(2.4) and (2.10), the random variables

$$z_h = \frac{\bar{y}_h - \bar{y}_0 - \nu_h}{\sqrt{2\sigma^2 / n}}$$

are distributed as standard normal for each $h = 1,\ 2,..., K$. The covariance of these random variables can be calculated as

$$Cov(z_i, z_j) = \frac{1}{2\sigma^2/n} Cov(\bar{y}_i - \bar{y}_0, \bar{y}_j - \bar{y}_0) = \frac{1}{2\sigma^2/n} \cdot Var(\bar{y}_0) = \frac{1}{2} \text{ for } i \neq j .$$

Then, two-sided simultaneous confidence intervals for $\nu_l = \mu_l - \mu_0$, $l = 1,2,...K$ under the assumption of (2.11) for the case of K active doses can be obtained as follows.

The T statistic is bounded by $|c|_K^\alpha$ if and only if $\max_{1 \le i \le j \le K} \frac{\left|\sum_{h=i}^{j} z_h\right|}{\sqrt{j-i+1}} \le |c|_K^\alpha$, *i.e.*, if and only if

$-|c|_K^\alpha \sqrt{j-i+1} \le \sum_{h=i}^{j} (\bar{y}_h - \bar{y}_0) - \sum_{h=i}^{j} \nu_h \le |c|_K^\alpha \sqrt{j-i+1}$, *for all* $1 \le i \le j \le K$, which can be written as

$$\sum_{h=i}^{j} (\bar{y}_h - \bar{y}_0) - |c|_K^\alpha \sqrt{j-i+1} \le \sum_{h=i}^{j} \nu_h \le \sum_{h=i}^{j} (\bar{y}_h - \bar{y}_0) + |c|_K^\alpha \sqrt{j-i+1}, \quad \textit{for all } 1 \le i \le j \le K .$$

Dividing the last inequality by $j-i+1 > 0$, we obtain

$$\frac{\sum_{h=i}^{j} (\bar{y}_h - \bar{y}_0)}{j-i+1} - \frac{|c|_K^\alpha}{\sqrt{j-i+1}} \le \frac{\sum_{h=i}^{j} \nu_h}{j-i+1} \le \frac{\sum_{h=i}^{j} (\bar{y}_h - \bar{y}_0)}{j-i+1} + \frac{|c|_K^\alpha}{\sqrt{j-i+1}}, \quad \textit{for all } 1 \le i \le j \le K .$$

Taking the maximum of the left hand side over all indexes less or equal to l, and the minimum of the right hand side over all indexes greater or equal to l, we obtain two inequalities:

$$\max_{1 \le p \le q \le l} \left(\frac{\sum_{h=p}^{q} (\bar{y}_h - \bar{y}_0)}{q-p+1} - \frac{|c|_K^\alpha}{\sqrt{q-p+1}} \right) \le \max_{1 \le p \le q \le l} \frac{\sum_{h=p}^{q} \nu_h}{q-p+1}, \quad l = 1,2,...,K \tag{2.15.1}$$

$$\min_{l \le r \le s \le K} \frac{\sum_{h=r}^{s} \nu_h}{s-r+1} \le \min_{l \le r \le s \le K} \left(\frac{\sum_{h=r}^{s} (\bar{y}_h - \bar{y}_0)}{s-r+1} - \frac{|c|_K^\alpha}{\sqrt{s-r+1}} \right), \quad l = 1,2,...,K . \tag{2.15.2}$$

Next, from (2.11) it follows that

$$\max_{1 \le p \le q \le l} \frac{\sum_{h=p}^{q} \nu_h}{q-p+1} = \nu_l, \quad \min_{l \le r \le s \le K} \frac{\sum_{h=r}^{s} \nu_h}{s-r+1} = \nu_l, \quad \textit{for all } l = 1,2,...,K ,$$

so the inequalities (2.15.1) and (2.15.2) can be combined to provide simultaneous confidence bands

$$\max_{1\le p\le q\le l}\left(\frac{\sum_{h=p}^{q}(\bar{y}_h-\bar{y}_0)}{q-p+1}-\frac{|c|_K^{\alpha}}{\sqrt{q-p+1}}\right)\le \nu_l \le \min_{l\le r\le s\le K}\left(\frac{\sum_{h=r}^{s}(\bar{y}_h-\bar{y}_0)}{s-r+1}-\frac{|c|_K^{\alpha}}{\sqrt{s-r+1}}\right),\ l=1,2,...,K\ . \quad (2.16)$$

2.3.3 Constructing Two-Sided Simultaneous Max-Min Confidence Bands for the General Case

The confidence bands above are constructed under the conditions of equal group sample sizes and equal known group variances. To generalize the bands to the case of arbitrary sample sizes and unknown (but common) variances, we omit assumption (2.10) and construct the procedure based on (2.1)-(2.4) only. Then $\bar{y}_i \sim N(\mu_i, \sigma^2/n_i)$, $i=0,\ 1,...,K$ are the sample means of the observations at each dose level i and $z_h=\dfrac{\bar{y}_i-\bar{y}_0-\nu_h}{\sqrt{\dfrac{\sigma^2}{n_i}+\dfrac{\sigma^2}{n_0}}}$ is distributed as standard normal for each $h=1,\ 2,...,K$. Since the group sample means are independent for $i=0,\ 1,...,K$, the covariance can be calculated as

$$Cov(z_i,z_j)=Cov\left(\frac{\bar{y}_i-\bar{y}_0}{\sqrt{\frac{\sigma^2}{n_i}+\frac{\sigma^2}{n_0}}},\frac{\bar{y}_j-\bar{y}_0}{\sqrt{\frac{\sigma^2}{n_j}+\frac{\sigma^2}{n_0}}}\right)$$

$$=Cov\left(\frac{-\bar{y}_0}{\sqrt{\frac{\sigma^2}{n_i}+\frac{\sigma^2}{n_0}}},\frac{-\bar{y}_0}{\sqrt{\frac{\sigma^2}{n_j}+\frac{\sigma^2}{n_0}}}\right)$$

$$=\frac{1}{\sqrt{\sigma^2\frac{n_0+n_i}{n_0n_i}}\sqrt{\sigma^2\frac{n_0+n_j}{n_0n_j}}}\cdot Var(\bar{y}_0)$$

$$= \frac{1}{\sqrt{\sigma^2 \frac{n_0 + n_i}{n_0 n_i}} \sqrt{\sigma^2 \frac{n_0 + n_j}{n_0 n_j}}} \cdot \frac{\sigma^2}{n_0}$$

$$= \sqrt{\frac{n_i n_j}{(n_0 + n_i)(n_0 + n_j)}} \quad \text{for } i \neq j .$$

Next, we compute the test statistic

$$T_{gen} = \max_{1 \le i \le j \le K} \frac{\left| \sum_{h=i}^{j} z_h \right|}{\sqrt{j - i + 1}} \Bigg/ \frac{\hat{\sigma}}{\sigma},$$

where $\nu \cdot \frac{\hat{\sigma}^2}{\sigma^2} \sim \chi^2_\nu$ with $\nu = \sum_{i=0}^{K} n_i - (K+1)$ and $\frac{\hat{\sigma}}{\sigma}$ is independent of z_h, $h = 1,2,...,K$. The critical values of the distribution of T_{gen}, $|c|^{\alpha}_{K,\nu}$, can be obtained by simulation for given α, K and ν. Then two-sided simultaneous confidence intervals for $\nu_l = \mu_l - \mu_0$, $l = 1,2,...,K$ can be obtained similarly as in the simple case with a few modifications. First, we consider the inequality $T_{gen} \le |c|^{\alpha}_{K,\nu}$ which has probability $1 - \alpha$ and is identical to

$$\max_{1 \le i \le j \le K} \frac{\left| \sum_{h=i}^{j} \bar{y}_h - \bar{y}_0 - \nu_h \right|}{\sqrt{\frac{\sigma^2}{n_h} + \frac{\sigma^2}{n_0}} \cdot \sqrt{j - i + 1}} \Bigg/ \frac{\hat{\sigma}}{\sigma} \le |c|^{\alpha}_{K,\nu}, \text{ or } \max_{1 \le i \le j \le K} \left| \sum_{h=i}^{j} \frac{\bar{y}_h - \bar{y}_0 - \nu_h}{\hat{\sigma} \cdot \sqrt{\frac{1}{n_h} + \frac{1}{n_0}} \cdot \sqrt{j - i + 1}} \right| \le |c|^{\alpha}_{K,\nu} .$$

This can be rewritten as

$$\left| \sum_{h=i}^{j} \frac{\bar{y}_h - \bar{y}_0 - \nu_h}{\hat{\sigma} \cdot \sqrt{\frac{1}{n_h} + \frac{1}{n_0}} \cdot \sqrt{j - i + 1}} \right| \le |c|^{\alpha}_{K,\nu}, \ \textit{for all } 1 \le i \le j \le K, \text{ or equivalently,}$$

$$\left| \sum_{h=i}^{j} \frac{\bar{y}_h - \bar{y}_0 - \nu_h}{\sqrt{\frac{1}{n_h} + \frac{1}{n_0}}} \right| \le |c|^{\alpha}_{K,\nu} \cdot \hat{\sigma} \cdot \sqrt{j - i + 1}, \ \textit{for all } 1 \le i \le j \le K, \text{ so that}$$

$$-|c|^{\alpha}_{K,\nu}\cdot\hat{\sigma}\cdot\sqrt{j-i+1}\le\sum_{h=i}^{j}\frac{\vec{y}_h-\bar{y}_0-\nu_h}{\sqrt{\frac{1}{n_h}+\frac{1}{n_0}}}\le|c|^{\alpha}_{K,\nu}\cdot\hat{\sigma}\cdot\sqrt{j-i+1},\ \textit{for all}\ 1\le i\le j\le K,$$ which

yields

$$\sum_{h=i}^{j}\frac{\bar{y}_h-\bar{y}_0}{\sqrt{\frac{1}{n_h}+\frac{1}{n_0}}}-|c|^{\alpha}_{K,\nu}\cdot\hat{\sigma}\cdot\sqrt{j-i+1}\le\sum_{h=i}^{j}\frac{\nu_h}{\sqrt{\frac{1}{n_h}+\frac{1}{n_0}}}\le\sum_{h=i}^{j}\frac{\bar{y}_h-\bar{y}_0}{\sqrt{\frac{1}{n_h}+\frac{1}{n_0}}}+|c|^{\alpha}_{K,\nu}\cdot\hat{\sigma}\cdot\sqrt{j-i+1}.\quad(2.17)$$

Next, we apply the following simple lemma.

Lemma 2.1. Let λ_h, $h=1,2,...,K$ be any non-negative constants. Then under monotonicity assumption (2.11),

$$\max_{1\le p\le q\le s}\frac{\sum_{h=p}^{q}\lambda_h\nu_h}{\sum_{h=p}^{q}\lambda_h}=\nu_s \text{ and } \min_{s\le p\le q\le K}\frac{\sum_{h=p}^{q}\lambda_h\nu_h}{\sum_{h=p}^{q}\lambda_h}=\nu_s, \text{ where } s=1,2,...,K.$$

Proof: Since $\nu_h\le\nu_s,\ \textit{for all}\ h\le s\le K$ and $\sum_{h=p}^{q}\lambda_h\nu_h\le\sum_{h=p}^{q}\lambda_h\nu_s=\nu_s\sum_{h=p}^{q}\lambda_h$,

for all $1\le p\le q\le s$ then

$$\frac{\sum_{h=p}^{q}\lambda_h\nu_h}{\sum_{h=p}^{q}\lambda_h}\le\nu_s\ \ \textit{for all}\ 1\le p\le q\le s.$$

The last inequality becomes an equality when $p=q=s$, which completes the proof of the first statement. The second statement can be proved similarly.

■

To use Lemma 2.1 for $\lambda_h^{-1}=\sqrt{\frac{1}{n_h}+\frac{1}{n_0}}$, we first divide (2.17) by $\sum_{h=i}^{j}\left(\frac{1}{n_h}+\frac{1}{n_0}\right)^{-1/2}$ and then take the maximum of the left hand side over all indexes less or equal to s, to obtain

$$\max_{1\le p\le q\le s}\left(\frac{\sum_{h=p}^{q}\frac{\bar{y}_h-\bar{y}_0}{\sqrt{\frac{1}{n_h}+\frac{1}{n_0}}}}{\sum_{h=p}^{q}\left(\sqrt{\frac{1}{n_h}+\frac{1}{n_0}}\right)^{-1}}-\left|c\right|_{K,\nu}^{\alpha}\cdot\hat{\sigma}\cdot\sqrt{q-p+1}\cdot\frac{1}{\sum_{h=p}^{q}\left(\sqrt{\frac{1}{n_h}+\frac{1}{n_0}}\right)^{-1}}\right)$$

$$\le \max_{1\le p\le q\le s}\frac{\sum_{h=p}^{q}\frac{\nu_h}{\sqrt{\frac{1}{n_h}+\frac{1}{n_0}}}}{\sum_{h=p}^{q}\left(\sqrt{\frac{1}{n_h}+\frac{1}{n_0}}\right)^{-1}}=\nu_s\,.$$

(2.18)

Similarly,

$$\nu_s=\min_{s\le p\le q\le K}\left(\frac{\sum_{h=p}^{q}\frac{\nu_h}{\sqrt{\frac{1}{n_h}+\frac{1}{n_0}}}}{\sum_{h=p}^{q}\left(\sqrt{\frac{1}{n_h}+\frac{1}{n_0}}\right)^{-1}}\right)$$

$$\le \min_{s\le p\le q\le K}\left(\frac{\sum_{h=p}^{q}\frac{\bar{y}_h-\bar{y}_0}{\sqrt{\frac{1}{n_h}+\frac{1}{n_0}}}}{\sum_{h=p}^{q}\left(\sqrt{\frac{1}{n_h}+\frac{1}{n_0}}\right)^{-1}}+\left|c\right|_{K,\nu}^{\alpha}\cdot\hat{\sigma}\cdot\sqrt{q-p+1}\cdot\frac{1}{\sum_{h=p}^{q}\left(\sqrt{\frac{1}{n_h}+\frac{1}{n_0}}\right)^{-1}}\right).$$

(2.19)

Finally, combining (2.18) and (2.19), we obtain the confidence bands for ν_s, $s=1,2,...,K$:

$$\max_{1\le p\le q\le s}\left(\frac{\sum_{h=p}^{q}\frac{\bar{y}_h-\bar{y}_0}{\sqrt{\frac{1}{n_h}+\frac{1}{n_0}}}}{\sum_{h=p}^{q}\left(\sqrt{\frac{1}{n_h}+\frac{1}{n_0}}\right)^{-1}}-\frac{|c|^{\alpha}_{K,\nu}\cdot\hat{\sigma}\cdot\sqrt{q-p+1}}{\sum_{h=p}^{q}\left(\sqrt{\frac{1}{n_h}+\frac{1}{n_0}}\right)^{-1}}\right)\le\nu_s$$

$$\le\min_{s\le p\le q\le K}\left(\frac{\sum_{h=p}^{q}\frac{\bar{y}_h-\bar{y}_0}{\sqrt{\frac{1}{n_h}+\frac{1}{n_0}}}}{\sum_{h=p}^{q}\left(\sqrt{\frac{1}{n_h}+\frac{1}{n_0}}\right)^{-1}}+\frac{|c|^{\alpha}_{K,\nu}\cdot\hat{\sigma}\cdot\sqrt{q-p+1}}{\sum_{h=p}^{q}\left(\sqrt{\frac{1}{n_h}+\frac{1}{n_0}}\right)^{-1}}\right). \tag{2.20}$$

Although we do not provide the formal proof that the two-sided intervals given by (2.16) and (2.20) are indeed the simultaneous confidence bands with $1-\alpha$ confidence level, the proof is similar to the one provided for the one-sided confidence bands, which is given as the proof of Lemma 2.2 in Section 2.4.1.

The confidence bands given by (2.16) and (2.20) can be used in clinical studies to examine the behavior of a test drug and to characterize the differences between the active treatment effects and the placebo effect. These bands are constructed under an assumption of a non-decreasing mean dose-response and can be modified for the case of a non-increasing function. Next, we illustrate how to construct the lower bands for the differences between responses to the active dose levels and placebo and discuss two procedures for identifying the minimum effective doses based on the lower bands.

2.4 MINIMUM EFFECTIVE DOSE DERIVED FROM ONE-SIDED CONFIDENCE BANDS

2.4.1 Constructing One-Sided Simultaneous Max-Min Confidence Bands

To construct one-sided confidence bands we consider the procedure similar to the one for obtaining the two-sided confidence intervals, but instead of the test statistics T and T_{gen} we use the statistics

$$T^0 = \max_{1\le i\le j\le k} \frac{\sum_{h=i}^{j} z_h}{\sqrt{j-i+1}} \tag{2.21}$$

for the case when assumption (2.10) holds and

$$T^0_{gen} = \max_{1\le i\le j\le k} \frac{\sum_{h=i}^{j} z_h}{\sqrt{j-i+1}} \Bigg/ \frac{\hat{\sigma}}{\sigma}, \tag{2.22}$$

for more general case of Section 2.3.3. The distributions of both statistics, T^0 and T^0_{gen} depend on the number k. For $k=1$, T^0 is distributed as a standard normal random variable and T^0_{gen} has the $t-$distribution with $\nu = n_0 + n_1 - 2$ degrees of freedom. The graph in Figure 3 illustrates the approximated distribution of T^0, obtained by simulation based on 100,000 replications for $k=10$ under the assumptions (2.2)-(2.4) and (2.10).

The critical values of both distribution functions can be also obtained by the simulation studies as the solutions $|m|^{\alpha}_{k}$ and $|m'|^{\alpha}_{k,\nu}$ to the equations $P(T^0 \le |m|^{\alpha}_{k}) = 1-\alpha$ and $P(T^0_{gen} \le |m'|^{\alpha}_{k,\nu}) = 1-\alpha$ for the specified values of α, k and ν. Table 2 contains the critical values of T_0 statistic (based on 100,000 replications) for up to ten active dose levels for $\alpha = 0.1$, $\alpha = 0.05$ and $\alpha = 0.01$.

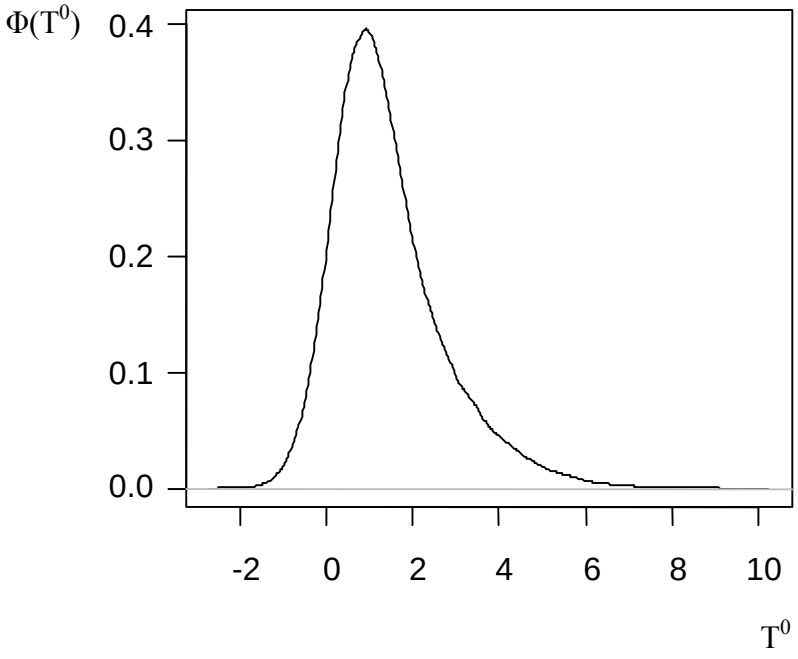

Figure 3. Approximated density function of T^0 statistic (2.21) for k =10

The simultaneous one-sided max-min confidence bands for mean response differences, $\nu_i's$, are given by

$$L_i = \max_{1\le p\le q\le i}\left(\frac{\sum_{h=p}^{q}(\bar{y}_h-\bar{y}_0)}{q-p+1}-\frac{|m|_k^{\alpha}}{\sqrt{q-p+1}}\right)\le \nu_i\ ,\ i=1,\ 2,...,k \tag{2.23}$$

based on the statistic T^0, and by

$$L_i'= \max_{1\le p\le q\le i}\left(\frac{\sum_{h=p}^{q}\frac{\bar{y}_h-\bar{y}_0}{\sqrt{\frac{1}{n_h}+\frac{1}{n_0}}}}{\sum_{h=p}^{q}\left(\sqrt{\frac{1}{n_h}+\frac{1}{n_0}}\right)^{-1}}-\frac{|m'|_{k,\nu}^{\alpha}\cdot\hat{\sigma}\cdot\sqrt{q-p+1}}{\sum_{h=p}^{q}\left(\sqrt{\frac{1}{n_h}+\frac{1}{n_0}}\right)^{-1}}\right)\le \nu_i,\ \ i=1,\ 2,...,k \tag{2.24}$$

based on the statistic T_{gen}^0. Next, we prove the following lemma.

Table 2. α -level critical values of T^0 statistic given by (2.21)

Number of Active	Significance Level		
Dose Levels	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$
K=1	1.28	1.64	2.33
K=2	1.72	2.12	2.89
K=3	2.01	2.45	3.34
K=4	2.24	2.73	3.72
K=5	2.42	2.96	4.06
K=6	2.59	3.19	4.38
K=7	2.76	3.39	4.71
K=8	2.89	3.59	4.99
K=9	3.04	3.79	5.23
K=10	3.17	3.94	5.50

Lemma 2.2. Let $C(\nu)$ be the confidence level of the max-min simultaneous confidence intervals (2.23) and (2.24). Then $C(\nu)$ is at least $1-\alpha$.

Proof: We prove the lemma for the confidence bands given by (2.23). Similar proof can be used for the bands given by (2.24). First, from Lemma 2.1 it follows that

$$\nu_i = \max_{1\le p\le q\le i} \frac{\sum_{h=p}^{q} \nu_h}{q-p+1}.$$

Then the confidence level is defined as

$$C(\nu) = P(L_i \le \nu_i, i = 1,2,...,k)$$

$$= P\left(\max_{1\le p\le q\le i} \left[\frac{\sum_{h=p}^{q} (\bar{y}_h - \bar{y}_0)}{q-p+1} - \frac{|m|_k^{\alpha}}{\sqrt{q-p+1}} \right] \le \max_{1\le p\le q\le i} \frac{\sum_{h=p}^{q} \nu_h}{q-p+1}, \quad for\ all\ i = 1,2,...,k \right)$$

$$= P\left(\max_{1\le p\le q\le i} \left[\frac{\sum_{h=p}^{q} (\bar{y}_h - \bar{y}_0)}{q-p+1} - \frac{|m|_k^{\alpha}}{\sqrt{q-p+1}} \right] - \max_{1\le p\le q\le i} \frac{\sum_{h=p}^{q} \nu_h}{q-p+1} \le 0, \quad for\ all\ i = 1,2,...,k \right)$$

$$\ge P\left(\max_{1\le p\le q\le i} \left[\frac{\sum_{h=p}^{q} (\bar{y}_h - \bar{y}_0) - (\mu_h - \mu_0)}{q-p+1} - \frac{|m|_k^{\alpha}}{\sqrt{q-p+1}} \right] \le 0, \quad for\ all\ i = 1,2,...,k \right)$$

$$= P\left(\max_{1\le p\le q\le i} \left[\frac{\sum_{h=p}^{q} z_h}{q-p+1} - \frac{|m|_k^{\alpha}}{\sqrt{q-p+1}} \right] \le 0, \quad for\ all\ i = 1,2,...,k \right)$$

$$= P\left(\left[\frac{\sum_{h=p}^{q} z_h}{q-p+1} - \frac{|m|_k^{\alpha}}{\sqrt{q-p+1}} \right] \le 0, \quad for\ all\ 1 \le p \le q \le i,\ i = 1,2,...,k \right)$$

$$= P\left(\frac{\sum_{h=p}^{q} z_h}{\sqrt{q-p+1}} \le |m|_k^{\alpha}, \quad for\ all\ 1 \le p \le q \le i,\ i = 1,2,...,k \right)$$

$$= P\left(\max_{1\le p\le q\le k} \frac{\sum_{h=p}^{q} z_h}{\sqrt{q-p+1}} \le |m|_k^{\alpha} \right)$$

$$= P(T^0 \le |m|_k^{\alpha}) = 1-\alpha,$$

which is a true statement, because the critical value is calculated as the solution to this equation for $|m|_k^{\alpha}$.

■

2.4.2 Procedure for Identifying the Minimum Effective Dose

Before we propose a closed step-down procedure based on the MPG principle, let us mention that the lower bands given by (2.23) and (2.24) can readily be used to provide the simultaneous lower confidence statements for the mean response differences at all dose levels. For any given number of active drug levels, K, and confidence level, $1-\alpha$, the positive values of the lower bands indicate effective doses of the drug and the smallest such dose can be used as an MED estimate. Doses where the lower bounds are negative indicate non-effective doses. Next, we construct a more powerful procedure for estimating the MED.

We propose a step-down procedure to test the hypotheses (2.13) based on constructing $1-\alpha$ simultaneous one-sided max-min confidence bands for mean response differences, ν_i's, given by (2.23) and (2.24). For simplicity, we illustrate the proposed step-down procedure using the intervals (2.23)

$$L_l = \max_{1\le p\le q\le l}\left(\frac{\sum_{h=p}^{q}(\bar{y}_h-\bar{y}_0)}{q-p+1}-\frac{|m|_k^{\alpha}}{\sqrt{q-p+1}}\right)\le \nu_l,\ l=1,\ 2,...,k,\ 1\le k\le K.$$

This method uses only the lower bands for the $l=k-th$ mean dose-response difference. Starting with $l=k=K$, H_{0K} is tested against H_{aK} at a level of significance α by computing L_K using the corresponding critical value $|m|_K^{\alpha}$, where H_{0K} and H_{aK} are given by (2.13). Positive L_K suggests that at least one of the K mean response differences is positive so the corresponding null hypothesis H_{0K} is rejected. Next, the highest dose is omitted from consideration and the first $k=K-1$ doses are considered and the lower bound for $\nu_{k=K-1}$ is calculated based on $|m|_{k=K-1}^{\alpha}$. If $L_k>0$ then $H_{0k=K-1}$ is rejected and the testing is continued with $k=K-2$ and so on unless $L_m\le 0$ for some $m=1,\ 2,...,K$, then H_{0m} is accepted and testing is completed. The MED is estimated as $MED^*=m+1$. If $m=K$ then $MED^*=K+1$, meaning that there are no effective doses in the study. If such $m=1,\ 2,...,K$ can not be found, *i.e.*, all hypotheses are rejected, then all of the active doses are estimated being effective and the smallest active dose corresponds to the minimum effective dose, *i.e.*, $MED^*=1$.

The above procedure is given for the closed family of hypotheses (2.13) (See Section 2.2) and uses the MPG testing principle. Indeed, each hypothesis is tested at a fixed level of significance and a hypothesis is tested and rejected if and only if it is significant and all hypotheses implying it are also significant. For example, the hypothesis H_{0m} for some $m < K$ is tested if and only if all hypotheses with higher indexes $H_{0m+1}, H_{0m+2}, ..., H_{0K}$ are rejected. Since the set of hypotheses $H_{0m+1}, H_{0m+2}, ..., H_{0K}$ contains all of the hypotheses implying H_{0m}, H_{0m} is tested if and only if all hypotheses implying it are significant. Moreover, H_{0m} is rejected if and only if it is significant along with all hypotheses implying it.

A simpler one-step alternative approach would be to construct the simultaneous lower bounds for L_l, $l = 1,2,...,k$ with $k = K$. Then starting with the highest dose difference $l = K$, if $L_l > 0$ then H_{0l} is rejected and H_{0l-1} is tested for some $l,\ 2 \le l \le K$. The maximum $l,\ 2 \le l \le K$ such that H_{0l-1} is accepted provides the estimated MED. If all lower bounds are positive then MED is estimated as the first dose and if $L_K \le 0$ then we conclude that the MED does not exist.

The proposed procedure is more powerful than the above alternative approach. The latter one uses a single critical value, $|m|_K^{\alpha}$, for constructing the bands. Since the critical value $|m|_K^{\alpha} > |m|_l^{\alpha}$ for $1 \le l < K$, the proposed procedure is more powerful. Note that the procedures perform similarly only in the case when the lower bound for the highest mean dose-response difference is negative, therefore, there are no effective doses detected by the procedures.

Example 2.1. To illustrate the proposed step-down procedure, let us consider hypothetical data given by Tamhane *et al.* (1996). These authors suppose that in a single drug trial with five active dose levels and a placebo the same number of subjects are allocated to each of the six groups. The sample group means are given by

$\bar{y}_0$	$\bar{y}_1$	$\bar{y}_2$	$\bar{y}_3$	$\bar{y}_4$	$\bar{y}_5$
0.0	1.5	2.1	1.9	2.3	2.1

They further assume that (2.10) holds, *i.e.*, $2\sigma^2 / n = 1$ with $\nu = \infty$ degrees of freedom. The problem of identifying the MED is considered subject to the condition that the FWE is

controlled at 0.05. Among the results of several procedures presented by Tamhane *et al.* (1996), the results of Williams' procedure (2.8) and the results of SDPC procedure based on (2.9) are presented. Both methods provide the estimate $MED^* = 2$.

Now we illustrate the proposed step-down procedure. Since the placebo sample group mean is zero, the confidence bounds (2.23) reduce to

$$L_l = \max_{1\le p\le q\le l}\left(\frac{\sum_{h=p}^{q}\bar{y}_h}{q-p+1} - \frac{|m|_k^{\alpha}}{\sqrt{q-p+1}}\right),\ l=1,\ 2,...,k,\ 1\le k\le 5.$$

Table 3. Calculation of L_5 for example of Section 2.4.2

p	q	S_1	q-p+1	S_2	S_3	S_4
1	1	1.50	1.00	1.50	2.96	-1.46
1	2	3.60	2.00	1.80	2.09	-0.29
1	3	5.50	3.00	1.83	1.71	0.12
1	4	7.80	4.00	1.95	1.48	0.47
1	5	9.90	5.00	1.98	1.32	0.66
2	2	2.10	1.00	2.10	2.96	-0.86
2	3	4.00	2.00	2.00	2.09	-0.09
2	4	6.30	3.00	2.10	1.71	0.39
2	5	8.40	4.00	2.10	1.48	0.62
3	3	1.90	1.00	1.90	2.96	-1.06
3	4	4.20	2.00	2.10	2.09	0.01
3	5	6.30	3.00	2.10	1.71	0.39
4	4	2.30	1.00	2.30	2.96	-0.66
4	5	4.40	2.00	2.20	2.09	0.11
5	5	2.10	1.00	2.10	2.96	-0.86

Max = 0.66

First, we need to calculate L_5, based on $|m|_5^{0.05} = 2.96$ (see Table 2). We start with computing $S_1 = \sum_{h=p}^{q}\bar{y}_h$ for all possible combinations of indexes p and q and then, we find

$S_2 = \dfrac{S_1}{q-p+1}$ and $S_3 = \dfrac{2.96}{\sqrt{q-p+1}}$ and finally, calculate $S_4 = S_2 - S_3$. The detailed calculations based on this notation are given in Table 3. Since $L_5 = 0.66 > 0$ we reject H_{05} and test H_{04}. Similarly, for $|m|_4^{0.05} = 2.73$ we obtain $L_4 = 0.59 > 0$ so we reject H_{04} and test H_{03} based on $|m|_3^{0.05} = 2.45$. $L_3 = 0.42 > 0$ so we reject H_{03} and test H_{02} with $|m|_2^{0.05} = 2.12$. Finally, $L_2 = 0.30 > 0$ and $L_1 = -0.14 \le 0$ based on $|m|_1^{0.05} = 1.64$. So we accept H_{01} and estimate $MED^* = 2$.

2.5 SIMULATION STUDY

Tamhane *et al.* (1996) consider a simulation study to compare the results of a number of procedures. For the monotone dose-response, the studies assume two types of population response functions: linear and step (the examples of such types of the response functions are given in Figure 1). We consider a similar set up for the simulation studies to compare the performance of our proposed procedure with the Williams' (2.8) and the SDPC (2.9) procedures. The critical values under the assumption (2.10) for the latter two test procedures are given by Tamhane *et al.* (1996).

We assume a single drug study with a placebo response fixed at zero and $K = 5$ active dose levels with the significance level $\alpha = 0.05$. Under the assumption (2.10) of known variance and equal group sample sizes, we consider two types of monotone (non-decreasing) dose-response curves, linear and step. Let δ denote the largest value of the response function, this value corresponds to the 5th dose level. Tamhane *et al.* (1996) uses $\delta = 3$ and $\delta = 5$; in addition to these values, we consider $\delta = 4$. For each combination of the type of response and its maximum value, we vary the true MED over the five dose levels. Note that for $MED = 5$ the linear and step functions provide the same values of means. For each case of given mean values, we generate the sample mean responses $\bar{y}_i \sim N(\mu_i,\ 1/2)$, $i = 0,\ 1,...,5$, $\mu_0 = 0$ and apply the proposed step-down procedure, together with Williams' (WILM) procedure and the step-down

procedure based on pairwise contrasts (SDPC) to estimate the MED. Based on 10,000 replications for each case, we obtain the proportion when each dose is estimated as the MED. Since we know a priori the true MED and what doses correspond to the effective and non-effective doses, we can simulate the values of FWE, Power, Lack of Power and Bias in each case, based on the formulas below, given by Tamhane *et al.* (1996).

$$FWE^* = \frac{N_1}{N},\ Power^* = \frac{N_2}{N},\ Lack\ \ of\ \ Power^* = \frac{N_3}{N},\ Bias^* = E(MED^*) - MED, \quad (2.25)$$

where N_1 is the number of non-effective doses estimated as the MED's; N_2 is the number of doses, correctly estimated as the MED's; N_3 is the number of doses, higher than the MED being estimated as the MED, including the cases, when none of the doses is estimated as the MED; $E(MED^*)$ is the expected value of the estimated MED, based on the observed probabilities; if none of the five doses is estimated as the MED then $MED^* = 6$ is used to calculate the expectation and $N = 10{,}000$ is the total number of replications.

Example 2.2. Suppose we consider the linear response with its maximum value $\delta = 5$ and the true $MED = 3$, then the population means are given by $\mu_0 = \mu_1 = \mu_2 = 0$, $\mu_3 = 5/3$. $\mu_4 = 10/3$ and $\mu_5 = 5$. Next, we generate $\bar{y}_i \sim N(\mu_i,\ 1/2)$, $i = 0,\ 1,...,5$ and apply each of the three procedures. Each one provides an MED-estimate, $MED^* \in \{1,\ 2,\ 3,\ 4,\ 5,\ 6\}$. If we repeat the above steps 10,000 times, then each procedure produces the estimated MED's with the corresponding frequencies. For example, if we let d_i denote the number of times when a certain procedure provides the estimate $MED^* = i$ for $i = 1,\ 2,\ 3,\ 4,\ 5,\ 6$, then $\sum_{i=1}^{6} d_i = 10{,}000$. In this example with $MED = 3$, we can calculate $N_1 = d_1 + d_2$, $N_2 = d_3$, $N_3 = d_4 + d_5 + d_6$, $E(MED^*) = \frac{1}{10{,}000}\sum_{i=1}^{6} i \cdot d_i$.

The R Code for the simulation studies is given in Appendix B. It contains the procedures for calculating all of the above measures. Table 4 summarizes the familywise error, power and bias of three procedures. Since the simulation results are based on 10,000 replications, the standard error of estimation is $\sqrt{0.5(0.5)/10{,}000} = 0.005$.

Note that in the case of true $MED = 1$, there are no non-effective doses, so there can be no FWE. We also note that the results of our simulation studies agree with the results presented by Tamhane *et al.* (1996) for WILM and SDPC for $\delta = 3$ and $\delta = 5$.

First, as the theory establishes, all three procedures strongly control the FWE at $\alpha = 0.05$. Indeed, in order to conclude that significance level would be higher than 0.05, the estimated FWE must be greater than $0.05 + 1.96\sqrt{(0.05 \cdot 0.95)/10{,}000} = 0.054$. The largest simulated value of FWE across all considered population configurations corresponds to the Williams' procedure for linear shape of the mean dose-response with $\delta = 3$. But since it equals $0.052 < 0.054$, we conclude that all procedures strongly control the FWE, which agrees with the theoretical results.

When the true MED is relatively large then WILM and SDPC seem to perform better in terms of power and bias than the proposed procedure. The power of the proposed procedure is 0.378, which is less than the power of WILM (0.492) and the power of SDPC (0.443).

The above ranking is confirmed by the corresponding values of bias. The average bias of the WILM procedure is 0.097, followed by 0.109 of SDPC and by 0.197 of the proposed procedure.

In the cases with the true $MED = 1$, the proposed procedure performs the best. The average power across such configurations equals 0.473 for the proposed procedure, followed by 0.451 for the SDPC procedure and by 0.437 for the WILM procedure. The average biases corresponding to such designs are 1.655 for the proposed procedure, followed by 1.756 for the SDPC procedure and by 1.676 for the WILM procedure.

Based on these results we can conclude that the proposed procedure performs the best when the true MED is low, which is the most common case in the dose-response studies. And our procedure performs worse than the Williams' procedure and the SDPC procedure when the true MED is high.

If we compare the results corresponding to the same shape of the mean dose-response function for different values of δ, then, obviously, all procedures perform better in the case of $\delta = 5$, because the larger the mean dose-response difference the easier it can be detected by a statistical procedure.

If we compare the results corresponding to the same value of δ but different shapes of the mean dose-response function, then the average power of the procedures is higher for the step

function. It can be explained by the flat shape of the step function with zero mean difference for the consecutive non-effective dose levels and a relatively large positive difference associated with the dose effectiveness.

The performed simulation studies are limited to the equal sample size case, and its conclusion may be different for unequal sample size case. Moreover, only two shapes of the dose-response function are considered so additional studies are suggested for more detailed investigation of the performance of the procedures.

Table 4. Familywise error, power and bias

Response function	True MED	Estimated FWE (upper entry), power (middle entry) and bias (lower entry) with standard error of 0.005		
		Williams'	Step-down with pairwise contrasts	Proposed
Linear $\boxed{\delta = 3}$	1	0.000	0.000	0.000
		0.062	0.069	0.087
		3.316	3.553	3.487
	2	0.026	0.029	0.034
		0.061	0.052	0.053
		2.541	2.814	2.863
	3	0.030	0.034	0.039
		0.092	0.059	0.051
		1.816	2.097	2.225
	4	0.042	0.042	0.049
		0.195	0.103	0.062
		1.032	1.285	1.448
	5	0.052	0.049	0.049
		0.599	0.415	0.193
		0.253	0.390	0.613
Linear $\boxed{\delta = 4}$	1	0.000	0.000	0.000
		0.096	0.102	0.121
		2.563	2.833	2.787
	2	0.025	0.026	0.034
		0.100	0.079	0.077
		1.942	2.257	2.329
	3	0.040	0.039	0.045
		0.155	0.103	0.078
		1.320	1.616	1.796
	4	0.042	0.043	0.046
		0.318	0.188	0.103
		0.705	0.938	1.187
	5	0.048	0.045	0.048
		0.813	0.681	0.407
		0.051	0.142	0.408

Table 4 *continued*

Response function	True MED	Estimated FWE (upper entry), power (middle entry) and bias (lower entry)		
		Williams'	Step-down with pairwise contrasts	Proposed
Linear $\delta = 5$	1	0.000	0.000	0.000
		0.134	0.134	0.154
		1.977	2.213	2.181
	2	0.035	0.038	0.043
		0.144	0.111	0.100
		1.460	1.730	1.824
	3	0.041	0.042	0.045
		0.232	0.149	0.108
		0.965	1.218	1.420
	4	0.045	0.043	0.047
		0.462	0.307	0.172
		0.456	0.644	0.879
	5	0.047	0.046	0.049
		0.915	0.857	0.664
		-0.052	-0.036	0.146
Step $\delta = 3$	1	0.000	0.000	0.000
		0.543	0.588	0.636
		1.637	1.446	1.139
	2	0.046	0.044	0.047
		0.496	0.489	0.449
		1.282	1.281	1.220
	3	0.050	0.049	0.050
		0.500	0.446	0.333
		0.949	1.083	1.241
	4	0.048	0.049	0.048
		0.536	0.426	0.258
		0.604	0.793	1.055

Table 4 *continued*

Response function	True MED	Estimated FWE (upper entry), power (middle entry) and bias (lower entry)		
		Williams'	Step-down with pairwise contrasts	Proposed
Step $\delta = 4$	1	0.000	0.000	0.000
		0.829	0.846	0.869
		0.478	0.416	0.286
	2	0.049	0.049	0.049
		0.776	0.753	0.715
		0.367	0.396	0.388
	3	0.049	0.047	0.049
		0.769	0.715	0.600
		0.282	0.378	0.504
	4	0.048	0.048	0.049
		0.786	0.691	0.504
		0.169	0.291	0.536
Step $\delta = 5$	1	0.000	0.000	0.000
		0.960	0.964	0.969
		0.087	0.075	0.048
	2	0.051	0.051	0.051
		0.906	0.894	0.869
		0.029	0.051	0.058
	3	0.051	0.051	0.050
		0.904	0.875	0.814
		0.001	0.038	0.101
	4	0.051	0.051	0.049
		0.907	0.863	0.748
		-0.026	0.000	0.142

3.0 COMBINATION DRUG STUDY FOUNDATIONS

3.1 INTRODUCTION

Combination products, also known as fixed-dose combinations are combinations of two or more active drugs produced in a single tablet. Usually, these active drugs come from different classes of drugs with the same therapeutic effect. For example, the drug VICOPROFEN® (Abbott) is indicated for a short-term management of acute pain. This drug is a fixed-dose combination of the opioid analgesic agent, hydrocodone bitartrate (7.5 mg) and the nonsteroidal anti-inflammatory agent, ibuprofen (200 mg).

There are many reasons why it is desirable to combine drugs. The main rationale for combining drugs is to produce a better treatment. As in the case of VICOPROFEN®, the component drugs, opioid and ibuprofen, have different mechanisms of action. When these drugs are combined, then the combination produces an analgestic effect, which is significantly greater than the effect produced by each component drug alone (Palangio *et al.* (2000)).

Another rationale for combining drugs is to simplify treatment regiments for patients. Wertheimer and Morrison (2002) mention that the combination products are especially beneficial in the treating of infectious diseases, when exact compliance is crucial. These authors emphasize that "partial adherence can lead to the development of drug-resistant strains and a threat to public health".

In order for a new combination drug to be approved for human use, the FDA requires demonstrating that "each component makes a contribution to the claimed effects and the dosage of each component (amount, frequency, duration) is such that the combination is safe and effective for a significant patient population requiring such concurrent therapy as defined in the labeling for the drug" (21 CFR 300.50). In order to satisfy these requirements, safety and efficacy studies are performed.

Efficacy studies are needed to assess the dose-response relationship and characterize the effectiveness of combination drugs. There are two primary goals of such studies. The first one is to obtain the information about all effective doses and the second one is to demonstrate that each component makes a contribution to the claimed effects, *i.e.*, to show that there is a gain in effect of combining drugs at certain fixed doses.

Our research is devoted to the analysis of efficacy of combination drug studies. We assume the most commonly occurring case, when both single drugs are *known to be effective at the considered doses*. So if we identify all combinations which produce a therapeutic effect higher than the effect produced by each component alone, then all such superior combinations are considered by regulatory agencies to be effective. Our primary research goal is to detect the "lowest" doses among such superior combinations. Selecting of the lowest combination is crucial especially when administering higher doses of the drug can cause serious side effects so it is beneficial to administer lower (safer) doses.

The problem stated above is different from the problem of identifying the effective combinations, because the effective combinations do not necessarily produce a greater effect than the single drugs alone so additional analysis is needed to satisfy the FDA regulations regarding combination efficacy. Because we assume both individual components are effective, by demonstrating the FDA requirements, we also show effectiveness of drug combination. In theory, if one were using a component drug which is not known to be effective then in addition to showing the combination is superior to each component, one would need to demonstrate that the combination is also superior to placebo. Combination drugs of the latter type are not very common, and as such are not covered by our methodology.

3.2 POPULATION PARAMETERS USED IN COMBINATION DRUG STUDIES

3.2.1 Design of Combination Drug Studies and Matrix Representation of Population Parameters

In a $K \times N$ combination drug trial, subjects are randomly allocated to one of $(K+1)\times(N+1)$ groups. Subjects of each group receive a combination (x_i, y_j) of drug A at dose level $i,\ i = 0,...,K$ and drug B at dose level $j,\ j = 0,...,N$, where $(0, y_j)$ and $(x_i, 0)$ are the $j-th$ dose of drug B, and the $i-th$ dose of drug A alone, respectively. We are assuming in our research that both component drugs are effective.

There is another way to code the drug doses. If instead of the actual amounts of the doses, the dose levels are used for identifying the combination, then a drug combination (x_i, y_j) can be recorded as the combination (i, j). For example, if the first active dose of drug A is 3mg and the second active dose of drug B is 4mg, then the combination $(x_1, y_2) = (3mg,\ 4mg)$ can be referred to as a combination $(1,\ 2)$. We are going to use both representations.

Let μ_{ij} be the population mean dose response to a drug combination (x_i, y_j), *i.e.*, the population therapeutic effect produced by the combination (x_i, y_j).

The expected gain from combining the $i-th$ active dose of drug A with the $j-th$ active dose of drug B, is defined as $\theta_{ij} = \min\{\mu_{ij} - \mu_{i0}, \mu_{ij} - \mu_{0j}\}$, $i = 1,...,K$ and $j = 1,...,N$. Hung *et al.* (1993) states that if this gain is positive, then the combination is considered to be superior in terms of greater effect over its individual components alone. And since both component drugs are assumed to be effective, we are able to claim since the gain is positive that this combination is efficacious. Let E be the set of all such efficacious doses, *i.e.*,

$$E = \{(x_i, y_j), i = 1,..., K, j = 1,..., N : \theta_{ij} > 0\}.$$

We use the matrix notation to represent the response structure of the $K \times N$ trial. The matrix of mean dose-responses is a $(K+1)\times(N+1)$ matrix where the $ij-th$ entry corresponding to the population mean response of the drug combination (x_i, y_j) is given by

$$R = \begin{pmatrix} \mu_{00} & \mu_{01} & \cdots & \mu_{0N} \\ \mu_{10} & \mu_{11} & \cdots & \mu_{1N} \\ \cdots & \cdots & \cdots & \cdots \\ \mu_{K1} & \mu_{K2} & \cdots & \mu_{KN} \end{pmatrix}$$

The first row and first column of the matrix correspond to the controls: drug B alone and drug A alone respectively, μ_{00} is the placebo mean response.

Next, let us construct two $K \times N$ matrices of partial effects.

$$\Theta_A = \begin{pmatrix} \mu_{11} - \mu_{01} & \mu_{12} - \mu_{02} & \cdots & \mu_{1N} - \mu_{0N} \\ \mu_{21} - \mu_{01} & \mu_{22} - \mu_{02} & \cdots & \mu_{2N} - \mu_{0N} \\ \cdots & \cdots & \cdots & \cdots \\ \mu_{K1} - \mu_{01} & \mu_{K2} - \mu_{02} & \cdots & \mu_{KN} - \mu_{0N} \end{pmatrix},$$

$$\Theta_B = \begin{pmatrix} \mu_{11} - \mu_{10} & \mu_{12} - \mu_{10} & \cdots & \mu_{1N} - \mu_{10} \\ \mu_{21} - \mu_{20} & \mu_{22} - \mu_{20} & \cdots & \mu_{2N} - \mu_{20} \\ \cdots & \cdots & \cdots & \cdots \\ \mu_{K1} - \mu_{K0} & \mu_{K2} - \mu_{K0} & \cdots & \mu_{KN} - \mu_{K0} \end{pmatrix}.$$

The partial effects matrix Θ_A is obtained by subtracting the 1st row of matrix R from the rest of the rows, so Θ_A stores the differences between the combination drug mean response and the mean response of drug B alone taken at the same dose levels, *i.e.*, the partial effects due to agent A. Similarly, the partial effects matrix Θ_B is obtained by subtracting the 1st column of matrix R from the rest of the columns and it stores the partial effects due to drug B.

By taking the minimum of two $ij-th$ elements of Θ_A and Θ_B, we obtain the element θ_{ij} of matrix of expected gains $\Theta = \{\theta_{ij}\}$, $i = 1,...,K$ and $j = 1,...,N$. Using the fact that $\theta_{ij} = \min(\mu_{ij} - \mu_{i0}, \mu_{ij} - \mu_{0j}) = \mu_{ij} + \min(-\mu_{i0}, -\mu_{0j}) = \mu_{ij} - \max(\mu_{i0}, \mu_{0j})$, where $i = 1,...,K$ and $j = 1,...,N$, we obtain the expected gains matrix

$$\Theta = \begin{pmatrix} \mu_{11} - \max(\mu_{01}, \mu_{10}) & \mu_{12} - \max(\mu_{10}, \mu_{02}) & \cdots & \mu_{1N} - \max(\mu_{10}, \mu_{0N}) \\ \mu_{21} - \max(\mu_{20}, \mu_{01}) & \mu_{22} - \max(\mu_{20}, \mu_{02}) & \cdots & \mu_{2N} - \max(\mu_{20}, \mu_{0N}) \\ \cdots & \cdots & \cdots & \cdots \\ \mu_{K1} - \max(\mu_{K0}, \mu_{01}) & \mu_{K2} - \max(\mu_{K0}, \mu_{02}) & \cdots & \mu_{KN} - \max(\mu_{K0}, \mu_{0N}) \end{pmatrix}.$$

The matrices Θ_A, Θ_B will be referred to as the partial effect matrices and Θ will be referred to as the gain matrix.

3.2.2 Definitions of Matrix Partial Order and Isotonic Responses

We start examining the properties of the above matrices by first making the assumption of isotonic responses with respect to the matrix partial order. If the doses of drug A and drug B are coded as i, $i = 0,...,K$ and j, $j = 0,...,N$ respectively, then the matrix partial order $\prec$ on the grid $\{(i, j), i = 0,...,K, j = 0,...,N\}$ is defined as $(i, j) \prec (i', j')$ if and only if $i \le i'$ and $j \le j'$. Then a $(K+1)\times(N+1)$ matrix $A = \{a_{ij}\}$ is said to be isotonic with respect to the matrix partial order if $(i, j) \prec (i', j')$ implies that $a_{ij} \le a_{i'j'}$ for all i, $i = 0,...,K$ and j, $j = 0,...,N$. The assumption that μ_{ij} are isotonic with respect to the matrix partial order states that for all i, $i = 0,...,K$ and j, $j = 0,...,N$ if $(i, j) \prec (i', j')$ then $\mu_{ij} \le \mu_{i'j'}$.

The last is equivalent to saying that both drugs A and B have the monotone responses when the level of the other drug is fixed, *i.e.*, $\mu_{i0} \le \mu_{i1} \le ... \le \mu_{iN}$ and $\mu_{0j} \le \mu_{1j} \le ... \le \mu_{Kj}$ for all $i = 0,...,K$ and $j = 0,...,N$.

Under the assumption of isotonic structure, the matrix R has the following properties:

1) all elements are non-decreasing in each column,
2) all elements are non-decreasing in each row.

Matrix Θ_A (Θ_B) satisfies only the first (the second) property and nothing can be said about matrix Θ, except for that it consists of non-negative elements. In order to guarantee that both properties hold for partial effect matrices, we need additional assumptions on the μ_{ij}'s.

For convenience, let us fix the dose level of drug A and consider all of the $N+1$ doses of drug B, then the mean response function can be represented by a univariate function $g_i = g_i(y_j) = \mu_{ij}$, $j = 0,...,N$ for each fixed drug A dose level i, $i = 0,...,K$. The matrix of partial effects Θ_A consists of non-decreasing in each row elements if and only if $\mu_{s1} - \mu_{01} \le \mu_{s2} - \mu_{02} \le ... \le \mu_{sN} - \mu_{0N}$ for any $s = 1,...,K$.

Since the means are isotonic, all of the above mean differences are bounded by zero from below. The above set of inequalities is identical to $0 \le \mu_{st} - \mu_{0t} \le \mu_{s,t+1} - \mu_{0,t+1}$ for all $s = 1,...,K$ and $t = 1,...,N-1$, which can be rewritten as $0 \le \mu_{0,t+1} - \mu_{0t} \le \mu_{s,t+1} - \mu_{st}$ or as $0 \le g_0(y_{t+1}) - g_0(y_t) \le g_s(y_{t+1}) - g_s(y_t)$ for all $s = 1,...,K$ and $t = 1,...,N-1$.

So the elements of the partial effect matrix Θ_A are non-decreasing in each row if and only if the non-negative difference between mean responses to two consecutive doses of drug B for any fixed dose level of drug A is at least as big as the non-negative difference between mean responses to the same consecutive doses of drug B taken alone. Hence, this condition guarantees that Θ_A satisfies both properties, 1) and 2).

Moreover, if $g_i(y_j)$, $j = 0,...,N$ can be assumed linear for each fixed drug A dose level i, $i = 0,...,K$, then the condition becomes $0 \le m_0 \le m_s$ for all $s = 1,...,K$, where m_i is the slope of g_i, $i = 0,...,K$. So the rate at which the linear mean response corresponding to any fixed active drug A dose level increases is at least the rate at which the linear mean response corresponding to the placebo of drug A increases.

In a similar manner one can obtain the necessary and sufficient conditions to guarantee that partial effect matrix Θ_B consists of non-decreasing elements in each column, that is the non-negative difference between mean responses to two consecutive doses of drug A for any fixed dose level of drug B must be at least as big as the non-negative difference between man responses to the same consecutive doses of drug A taken alone. And in addition, if the mean response function to the combination is linear for each fixed dose level of drug B, then it is equivalent to the requirement that the linear mean response function for any fixed active dose of drug B increases at least as fast as the linear mean response corresponding to the placebo of A. Next, we are going to examine the structure of matrix Θ under the assumption that each matrix of partial effects has an isotonic structure.

Result 3.1. If the partial effect matrices Θ_A and Θ_B are isotonic then the matrix of gains Θ is isotonic.

Proof: Let us consider an arbitrary row, say the $k-th$ row, $1 \le k \le K$ of matrix Θ. If both matrices of partial effects consist of non-decreasing elements in each row and each column, then $0 \le \min(\mu_{k1} - \mu_{k0}, \mu_{k1} - \mu_{01}) \le \min(\mu_{k2} - \mu_{k0}, \mu_{k2} - \mu_{02}) \le ... \le \min(\mu_{kN} - \mu_{k0}, \mu_{kN} - \mu_{0N})$,

i.e., $0 \le \theta_{k1} \le \theta_{k2} \le ... \le \theta_{kN}$. Since we fix an arbitrary row index, the last fact is true for all $k = 1,..., K$, which proves that Θ is non-decreasing in each row.

Similarly, it can be shown that Θ is non-decreasing in each column. So the matrix of gains consists of non-negative elements and has an isotonic structure.

■

The following example illustrates that the converse to the statement is not true. Consider the matrix of mean responses R given by

$$R = \begin{pmatrix} \times & 0.0 & 0.0 \\ 0.1 & 0.2 & 1.1 \\ 0.0 & 0.4 & 1.0 \end{pmatrix},$$

where "×" indicates that this dose is not evaluated.
So that

$$\Theta = \Theta_B = \begin{pmatrix} 0.1 & 1.0 \\ 0.4 & 1.0 \end{pmatrix} \text{ and } \Theta_A = \begin{pmatrix} 0.2 & 1.1 \\ 0.4 & 1.0 \end{pmatrix}.$$

Then, obviously, matrix Θ is isotonic but Θ_A is not.

In the case of an isotonic non-negative matrix Θ, the behavior of the combination drug under investigation has a number of nice properties. First, since the gains are non-decreasing in each row and each column, each combination effect is at least as big as the corresponding effect of a single drug. Next, the combination expected gain non-decreases with increasing dose levels. The other advantages of an isotonic structure of the gain matrix are discussed later on, when the notion of the minimum efficacious combination is introduced.

3.2.3 Pharmaceutical Meaning of Negative and Zero Expected Gains

In the previous section we made some assumptions about the population dose-responses. Without these assumptions, the matrix Θ has a very general structure. This fact is illustrated by the following example. Suppose there are two active dose levels of each drug A and B and the population gains are $\theta_{11} \le 0,\ \theta_{12} > 0,\ \theta_{21} > 0$ and $\theta_{22} \le 0,$. In this case superiority of a dose compound (x_1, y_2) or (x_2, y_1) does not imply the superiority of the higher dose combination

(x_2, y_2). The reason for this is that the single drug effects are not necessarily synergistic or additive, *i.e.*, the effect of one single drug is not enhanced by another single drug taken at the given doses. This can be the consequence of pharmacodynamic and/or pharmacokinetic interaction. So for such drugs the assumption of the isotonic structure of the indicator gain matrix does not hold. But since, usually, a number of the single drug properties are established before the combination drug trial is conducted; some additional assumptions about the drug behavior can be made. In our research from now on we assume that all combination mean responses are at least as large as each of the single drug mean response taken at the same dose levels, *i.e.*, all $\theta_{ij} \geq 0$ with $\theta_{ij} = 0$ corresponding to dose combination with no gain. As it is discussed in literature review, given in Section 4.1, Hung *et al.* (1993) make the same assumption of non-negative gains.

3.2.4 Isotonic and General Structures of the Indicator Gain Matrix

For convenience, together with matrix Θ we use the "indicator gain matrix" Θ' which is $K \times N$ and has entries of 0 and 1 only. For any given gain matrix Θ, the elements of the indicator gain matrix Θ' are defined as

$$\theta'_{ij} = \begin{cases} 0, & if \ \ \theta_{ij} = 0 \\ 1, & if \ \ \theta_{ij} > 0. \end{cases}$$

There are two cases we are interested in examining. In the first case, we assume that the indicator gain matrix has an isotonic structure. The second case corresponds to a general structure of the indicator gain matrix.

In the isotonic case, the elements of each row and each column of Θ' create non-decreasing sequences, *i.e.*, Θ' has the property: if $\theta'_{ij} = 1$ for some $i = 1,..., K$ and $j = 1,..., N$, then $\theta'_{rs} = 1$ for all $r > i$ *and* $s > j$:

$$\Theta' = \begin{pmatrix} \text{.....} \\ 0\ 0\ 0\ \text{...}\ 0\ 0\ 0\ \text{...}0 \\ 0\ 0\ 0\ \text{....}0\ 1_{ij}\ 1\ \text{...}\ 1 \\ \text{.....} \\ 0\ 0\ 0\ \text{....}0\ 1\ \ 1\ \text{...}\ 1 \\ \text{.....} \end{pmatrix}.$$

Such structure of the indicator gain matrix implies that if there is a positive gain combination, then all combinations with higher row or column indexes are also more beneficial than the corresponding single components. Note that one of the sufficient conditions for this to hold is when the matrix Θ itself is isotonic.

In the general case, $\theta'_{ij} = 1$ for some $i = 1,...,K$ and $j = 1,...,N$ does not imply $\theta'_{rs} = 1$ for $r > i$ *and* $s > j$ and Θ' could look like

$$\Theta' = \begin{pmatrix} \text{.....} \\ 0\ \text{....}0\ 1_{ij}\ \text{...}0\ \text{...}1\ \text{...} \\ \text{.....} \\ \text{.....}\ 0\ 0\ \text{.....}1\ \text{...}0\text{...} \\ \text{.....} \end{pmatrix}.$$

Next, we are going to introduce the notion of the minimum efficacious dose in the combination drug study and examine its properties with regard to isotonic and general cases of the indicator gain matrix.

3.2.5 Definition of the Minimum Efficacious Combination, Non-Uniqueness and Definition of the MeD-Set

We define the minimum efficacious dose (MeD) using the Food and Drug Administration's policy (21 CFR 300.50) which states that "two or more drugs may be combined in a single dosage form when each component makes a contribution to the claimed effects." We appear to be the first to consider minimum efficacious doses in terms of the lowest combinations that produce an effect greater than each of the single drugs taken alone. We note that in those rare instances when the components themselves are not effective in the dose range being studied,

establishing that the combination is superior to each component while necessary under FDA regulation (21 CFR 300.50) is not sufficient to establish efficacy. That would require, in addition, showing superiority to placebo. In these special instances our minimum efficacious dose notion would need to be reinterpreted as finding the minimum doses which produce positive gain.

Definition of the MeD. A positive gain dose, such that decreasing any agent dose leads to a no-gain compound is called the minimum efficacious dose (MeD), *i.e.*, the dose (x_r, y_s) for some $1 \le r \le K,\ 1 \le s \le N$ is the MeD if all of the following are true:

$$\theta_{rs} > 0;$$
$$\theta_{ij} = \theta_{rj} = \theta_{is} = 0, \textit{ for all } i < r, j < s.$$

The set of all such doses is called the MeD-set. We introduce the notion of the *MeD-set*, because the MeD is not necessarily unique. Non-uniqueness can be illustrated by considering the example of Section 3.2.3 for the 2×2 case. There are 2 active dose levels of each drug and the population expected gains are $\theta_{11} \le 0$, $\theta_{12} > 0$, $\theta_{21} > 0$, $\theta_{22} > 0$, then (x_1, y_2) and (x_2, y_1) are both the MeD's, because both of these drug combinations produce a positive expected gain and the combination (x_1, y_1) is a no-gain combination..

Although, the above definition of the MeD given for the case when all $\theta_{ij} \ge 0$, it can be easily generalized when this assumption does not hold by replacing all equality signs with the "$\le$" sign and rewriting the last equality as three inequalities.

3.3 POPULATION MeD-SET

3.3.1 Matrix Representation and Properties of the Population MeD-Set

Since by definition, the MeD-set is the collection of all the efficacious doses, such that decreasing of at least one agent dose leads to a no-gain compound, the matrix representation of the MeD-set can be easily obtained from Θ or Θ'. The MeD-matrix is defined as an $K \times N$ matrix $\Lambda = \{\lambda_{ij}\}$ where for $i = 1,...K,\ \ j = 1,..., N$

$$\lambda_{ij} = \begin{cases} 0, \ \textit{if} \ \ \theta'_{ij} = 0 \\ 1, \ \textit{if} \ \ \theta'_{rs} = 0 \ \ \textit{for} \ \ r \le i, \ \ s \le j, \ \ \textit{except} \ \ \textit{for} \ \ \theta'_{ij} = 1. \end{cases}$$

There is a one-to-one correspondence between the elements of the MeD-set and the elements of Λ: $(x_i, y_j) \in MeD - set$ if and only if $\lambda_{ij} = 1$. So the properties of the population MeD-set can be described in terms of the MeD-matrix Λ.

From the definition of $\Lambda = \{\lambda_{ij}\}$ it follows that each row and each column of this matrix has at most a single entry of 1, hence, there is at most one MeD for any fixed dose level of drug A and at most one MeD for each fixed dose level of drug B. For the $K \times N$ design, the MeD-set can be empty; it can contain a single element; two elements; *etc.*, up to $\min\{K, N\}$ elements. For each fixed cardinality of the MeD-set there is a collection of different MeD-sets. The cardinalities of these collections are calculated later on. For example, in the 2×2 case, the following are all possible MeD-matrices:

$$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \text{ and } \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \qquad (3.1)$$

Here, the zero-matrix corresponds to the case when there are no efficacious doses and therefore, the MeD-set is empty. The matrices with a single non-zero entry identify the cases when the MeD is unique. And therefore, there are four possible MeD-sets of cardinality 1. The maximum number of MeD's is $\min\{2,2\} = 2$ and is represented by the last matrix in (3.1). The MeD-set of the example of Section 3.2.3 also corresponds to the last matrix with two MeD's.

Next, we present two results which illustrate some basic relationship between the MeD-matrix and the corresponding indicator gain matrix.

Result 3.2. For any MeD-matrix there exists a unique isotonic indicator gain matrix.

Proof: If Λ is a zero matrix, then there are no efficacious doses among the considered combinations and the corresponding indicator matrix is the zero matrix, which is isotonic. For non-zero Λ we construct the corresponding indicator gain matrix using the following approach. First, form a matrix $A = \Lambda$, then change all elements to the right and below entries of 1 from zero to one, *i.e.*, if $a_{st} = 1$, for some $1 \le s \le K$ and $1 \le t \le N$ then make $a_{ij} = 1$ for all $i = s, s+1, ..., K$ and $j = t, t+1, ..., N$. The final matrix A has a property that if an element is equal to 1, then all elements with the higher indexes are also 1. So it is isotonic and since all

combinations higher than the MeD's correspond to the efficacious doses (under the assumption of isotonic structure), this is the indicator gain matrix and it is unique by construction.

■

If the indicator gain matrix has a general structure, then superiority of some combinations does not imply superiority of combinations with higher dose levels, which is stated in the following result.

Result 3.3. If a $K \times N$ MeD-matrix $\Lambda = \{\lambda_{ij}\}$, satisfied $\lambda_{st} = 1$ where $s + t \le K + N - 2$, then there exist at least three different non-isotonic indicator gain matrices corresponding to the matrix $\Lambda = \{\lambda_{ij}\}$.

Proof: The condition stating that indexes of all non-zero elements λ_{st} satisfy $s + t \le K + N - 2$ guarantees that $\lambda_{K-1,N} = \lambda_{K,N-1} = \lambda_{K,N} = 0$. So the indicator gain matrices $\Theta'_t = \{\theta'^t_{ij}\}$, $t = 1,2,3$ can be constructed as follows. First, let $\theta'^t_{ij} = \lambda_{ij}$, for all i and j such that $i + j \le K + N - 2$ for all $t = 1,2,3$. The remaining three elements of each matrix $\Theta'_t = \{\theta'^t_{ij}\}$ can be specified using the assignments: $\theta'^1_{K-1,N} = \theta'^1_{K,N-1} = 1$, $\theta'^1_{K,N} = 0$, $\theta'^2_{K-1,N} = 0$, $\theta'^2_{K,N-1} = 1$, $\theta'^2_{K,N} = 0$ and $\theta'^3_{K-1,N} = 1$, $\theta'^3_{K,N-1} = \theta'^3_{K,N} = 0$. Then the matrices Θ'_1, Θ'_2 and Θ'_3 are non-isotonic and each of them corresponds to the given matrix $\Lambda = \{\lambda_{ij}\}$.

■

Example 3.1. To illustrate Results 3.2 and 3.3 let us consider a 3×3 case and suppose that the MeD-matrix is given by

$$\Lambda = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Then, we can claim that the combinations (1,1), (1,2), (2,1), (3,1) are not efficacious and the combinations (1,3), (2,2) are the MeD's, so they are superior. This provides the following information about the indicator gain matrix:

$$\Theta' = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & * \\ 0 & * & * \end{pmatrix}.$$

Next, in the case of the isotonic structure, the last matrix becomes

$$\Theta' = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix},$$

but in a general case the elements labeled by " * " stay unspecified and can be either 0 or 1. There are eight different possible assignments of 0 and 1 to these three unspecified elements and only one of the assignments corresponds to the isotonic indicator gain matrix, which is given above. All other assignments (seven of them) produce non-isotonic indicator gain matrices.

3.3.2 Recursive Algorithm for Identifying the Population MeD-Set for a Given Response Matrix

Using the definition of the MeD-set and the properties of its MeD-matrix Λ, we now provide an algorithm for identifying the minimum efficacious doses for the $K \times N$ case. We assume that the mean response matrix R is given, so the matrices Θ and Θ' can be constructed.

Algorithm

Step 1: Construct the indicator gain matrix Θ' from the given mean response matrix R. The $K \times N$ matrix Θ' has entries of $0's$ and $1's$ only.

Step 2 "Identifying the MeD of the combination drug for each fixed dose level of drug A": construct a new $K \times N$ matrix F with elements $\{f_{ij}\}$ defined as

$$f_{ij} = \begin{cases} 0, \ \textit{if} \ \theta'_{ij} = 0 \ \textit{or} \ \{\theta'_{ij} = 1 \ \textit{and} \ \exists s < j : \theta'_{is} = 1\} \\ 1, \ \textit{if} \ \theta'_{ij} = 1 \ \textit{and} \ \theta'_{is} = 0 \ \textit{for all} \ s < j. \end{cases}$$

Thus, matrix F has at most one non-zero entry in each row; that non-zero entry corresponds to the MeD of the combination drug for a fixed dose level of drug A. Let $(l_1, \ l_2)$ be the location of the first (from the left) non-zero element of the first (from the top) non-zero row of matrix F. If all rows of the matrix are zero, then stop and conclude that the MeD-set is empty, otherwise go to Step 3.

Step 3 "Identifying the MeD of the combination drug for each fixed dose level of drug B": Matrix F has at most a single non-zero entry in each row but it might have several non-zero

entries in a column, so the matrix must be modified (to a new matrix F') so that each column and each row has at most one non-zero column. Define F' with elements $\{f_{ij}'\}$ given by

$$f_{ij}' = \begin{cases} 0, \ if \ \ f_{ij} = 0 \ \ or \ \ \{f_{ij} = 1 \ \ and \ \ \exists r < s : f_{rj} = 1\} \\ 1, \ if \ \ f_{ij} = 1 \ \ and \ \ f_{rj} = 0 \ \ for \ \ all \ \ r < i. \end{cases}$$

Note that the non-zero entries of matrix F' represent only those efficacious combinations that become no-gain combinations if the dose level of any one of the drugs is decreased. However, it is possible to obtain an efficacious combination from some of those by decreasing both indexes. Therefore, not all of the non-zero entries of matrix F' represent the MeD's. The next step eliminates the entries that do not represent the MeD's and constructs the MeD-matrix. Let $(m_1, \ m_2)$ be the location of the non-zero entry of the 1st non-zero column of matrix F' and go to Step 4.

Step 4 "Recursion": This step is recursive. For the first run, matrix F' is the working matrix; the working matrix W for a next run is calculated in the previous run. If the working matrix W has no non-zero entries, the recursion is over and the MeD-set obtained in the previous step is returned as the resulting one.

Let W be a non-zero working matrix of size $A \times B$ (for the first run, $A = K, \ B = N$) with elements $w_{ij}'s$. Let $(l_1, \ l_2)$ be the location of the non-zero entry of the first non-zero row of the working matrix and $(m_1, \ m_2)$ be the location of the non-zero entry of the first non-zero column of the working matrix, *i.e.*,

$$\begin{aligned} &w_{i,j} = 0 \ \text{ for } \ \forall i < l_1 \ \text{ and } \ \forall j \\ &w_{i,j} = 0 \ \text{ for } \ \forall i \ \text{ and } \ \forall j < m_2 \\ &w_{l_1,l_2} = 1 \\ &w_{m_1,m_2} = 1. \end{aligned}$$

Note that both $(l_1, \ l_2)$ and $(m_1, \ m_2)$ belong to the MeD-set. Moreover, if any other MeD's exist, then they have the first index in the interval $m_1 < i < l_1$ and the second index in the interval $l_2 < j < m_2$. Thus, if $m_1 - l_1 < 2$ or $l_2 - m_2 < 2$ then the recursion is finished, and the additional elements should be included in the set. If $(l_1, \ l_2) \neq (m_1, \ m_2)$ then two elements,

$(l_1,\ l_2)$ and $(m_1,\ m_2)$, to be included in the set. If $(l_1,\ l_2) = (m_1,\ m_2)$ then a single element is to be included. For the first run, these are the only MeD's included in the MeD-set.

If $m_1 - l_1 > 1$ and $l_2 - m_2 > 1$, then more MeD's might exist in the sub-matrix of dimension $(m_1 - l_1 - 1) \times (l_2 - m_2 - 1)$ starting at element w_{l_1+1, m_2+1} of matrix W. Let this sub-matrix be W and submit the sub-matrix for the next run of the recursion.

The elements of the resulting MeD-set identify the non-zero entries of the MeD-matrix; all other entries of the MeD-matrix must be set to zero.

■

The R-Code and an example of the resulting MeD-matrices are given in Appendix C.

3.3.3 The Cardinality of the MeD-Set and Number of Different MeD-Sets of Fixed Cardinality for the K by N Design

In the $K \times N$ setting, the number of elements (cardinality) of an MeD-set is bounded between zero (when the MeD-set is empty) and $\min\{K, N\}$. Let p be the cardinality of an MeD-set and $\{M_1,\ M_2, ..., M_q\}$ denote the collection of all different MeD-sets each having p elements. We show that q, the number of elements in $\{M_1,\ M_2, ..., M_q\}$ is a function of p, K and N and obtain the recursive formula for its calculation.

Let us start with $p = 1$, then $q(1, K, N) = K \cdot N$, which corresponds to the number of elements of the MeD-matrix. For example, for the 2×3 design, there are six different MeD-sets of a single element $\{(1,1)\}$, $\{(1,2)\}$, $\{(1,3)\}$, $\{(2,1)\}$, $\{(2,2)\}$ and $\{(2,3)\}$.

Next, we consider $p = 2$, then the MeD-set can be written in the form $\{(i_1, j_1),\ (i_2, j_2)\}$. By the definition of the MeD, we must have $i_1 < i_2$ and $j_1 > j_2$. The first element with matrix coordinates (i_1, j_1) can be located anywhere except for the last row and the first column of the MeD-matrix, *i.e.*, $i_1 = 1,2,...K-1$ and $j_1 = 2,3,..., N$. The second element, with coordinates (i_2, j_2) must be located to the left and down from the first element, *i.e.*, $i_2 = i_1 + 1,..., K$ and $j_2 = 1,2,..., j_1 - 1$. So there are $(K - i_1) \cdot (j_1 - 1)$ choices for the second element. Hence, the total number of different choices of such two elements is given by

$$q(2,K,N)=\sum_{i=1}^{K-1}\sum_{j=2}^{N}(K-i)(j-1)=\sum_{i=1}^{K-1}(K-i)\sum_{j=2}^{N}(j-1)=\frac{N(N-1)}{2}\sum_{i=1}^{K-1}(K-i)=\frac{K(K-1)N(N-1)}{4}$$

For example, for the 2×3 design, the MeD-sets of two elements are given by $\{(1,2),\ (2,1)\}$, $\{(1,3),\ (2,1)\}$ and $\{(1,3),\ (2,2)\}$ which agrees with $q(2,2,3)=\frac{2(1)3(2)}{4}=3$.

For the case of $p=3$, the MeD-set can be written in the form $\{(i_1,j_1),\ (i_2,j_2),\ (i_3,j_3)\}$, where $i_1<i_2<i_3$ and $j_1>j_2>j_3$. Here, $i_1=1,2,...K-2$ and $j_1=3,4,...,N$. The second and third elements must be located in a matrix with $K-i_1$ rows and j_1-1 columns. There are $q(2,K-i_1,j_1-1)$ different ways to choose the second and the third elements given the first element. So the total number of different choices of all three elements is given by

$$q(3,K,N)=\sum_{i=1}^{K-2}\sum_{j=3}^{N}q(2,K-i,j-1)=\sum_{i=1}^{K-2}\sum_{j=3}^{N}\frac{(K-i)(K-i-1)(j-1)(j-2)}{4}$$

So in a general case, the result can be summarized by the following theorem.

Theorem 3.1. In the $K\times N$ case, for any $p:\ 1\le p\le \min\{K,N\}$, the number of different MeD-sets of cardinality p, is given by the recursive formula

$$q(p,K,N)=\sum_{i=1}^{K-p+1}\sum_{j=p}^{N}q(p-1,K-i,j-1)\ \text{ for }\ 2\le p\le \min\{K,N\}\ \text{ with }\ q(1,K,N)=K\cdot N \qquad (3.2)$$

The above formula is very helpful if all MeD-sets must be identified for the given design, as it provides the total number of different MeD-sets containing the certain number of elements.

4.0 HYPOTHESES TESTING APPROACH TO ESTIMATE THE MeD-SET

4.1 LITERATURE REVIEW

We have examined the properties of the primary population parameters which are present in a combination drug trial. However, as we later note additional parameters also must be considered in order to draw some conclusions about the combination drug behavior.

Let y_{ijt} be the response of the $t-th$ subject in the $(i, j)-th$ dose combination group. We are going to introduce the procedure for identifying the MeD-set. This procedure is based on the following distributional properties and assumptions:

Assumption 1. Group sample sizes are equal $n_{ij} = n$, for $i = 0,...,K$ and $j = 0,...,N$,

(4.1a)

Assumption 2. The group sample mean responses are independent and normally distributed with the common variance, $\bar{y}_{ij} \sim N(\mu_{ij}, \sigma^2 / n)$ for $i = 0,...,K$ and $j = 0,...,N$.

(4.1b)

The U.S. Food and Drug Administration's policy (21 CFR 300.50) regarding the use of a fixed-dose combination drug requires that each component makes a contribution to the claimed effect of the combination. In view of this requirement, a lot of research has been done to study the superiority of a combination drug.

Laska and Meisner (1989) consider the problem of identifying whether the combination treatment is better than each of the single component agents of a single fixed-dose combination (i, j) using a hypothesis testing approach. They consider the null hypothesis $H_0 : \mu_{ij} \le \mu_{i0}$ *or* $\mu_{ij} \le \mu_{oj}$ versus the alternative $H_1 : \mu_{ij} > \mu_{i0}$ *and* $\mu_{ij} > \mu_{oj}$. They illustrate that an α-level test can be constructed by performing two separate statistical tests of $H_0^1 : \mu_{ij} \le \mu_{i0}$ versus $H_1^1 : \mu_{ij} > \mu_{i0}$ and $H_0^2 : \mu_{ij} \le \mu_{oj}$ versus $H_1^2 : \mu_{ij} > \mu_{oj}$ each at the same

α -level. Under the assumptions (4.1a) and (4.1b), the proposed test statistics are given by $S^1 = \sqrt{n/2} \cdot (\bar{y}_{ij} - \bar{y}_{i0})$ for testing H_0^1 and by $S^2 = \sqrt{n/2} \cdot (\bar{y}_{ij} - \bar{y}_{0j})$ for testing H_0^2 for the case when the variance is $\sigma^2 = 1$. The test statistics are jointly distributed as a bivariate normal random variable with mean vector $\sqrt{n/2}(\mu_{ij} - \mu_{i0}, \mu_{ij} - \mu_{oj})'$ and covariance matrix

$$\begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}.$$

Then the hypothesis H_0 is tested against H_1 by the "min-test", which rejects the null hypothesis if $\min\{S^1,\ S^2\} > c_\alpha$, where c_α is the upper α -level critical value of the standard normal distribution. In the normal case the min-test is a likelihood ratio test. Laska and Meisner also show that among the family of test statistics that are monotone non-decreasing functions of the two separate tests, the min-test is uniformly most powerful.

Hung *et al.* (1993) consider a more general problem of "global superiority". They consider a global null hypothesis associated with all active study combinations stating that all combinations in the study are no-gain combinations. The alternative hypothesis states that there exists at least one dose combination which is more effective than its respective component doses. Under the assumptions (4.1a) and (4.1b) two test statistics are constructed to test these hypotheses. The *AVE* test statistic, T_A, is given by

$$T_A = \sum_{i=1}^{K} \sum_{j=1}^{N} \hat{\theta}_{ij} \Big/ K \cdot N \cdot \hat{\sigma}\ , \tag{4.2}$$

where $\hat{\theta}_{ij}$, $\hat{\theta}_{ij} = \bar{y}_{ij} - \max(\bar{y}_{i0}, \bar{y}_{0j})$, is an estimator of θ_{ij}, the expected gain (or gain), for the $(i, j) - th$ dose combination. The pooled estimator $\hat{\sigma}^2$ of σ^2 is given by

$$\hat{\sigma}^2 = \sum_{i=0}^{K} \sum_{j=0}^{N} \sum_{t=1}^{n} \left(y_{ijt} - \bar{y}_{ij}\right)^2 \Big/ \left[(K+1) \cdot (N+1) \cdot (n-1)\right].$$

The second test statistic is the *MAX* statistic, T_{MAX}. It is given by

$$T_{MAX} = MAX(\hat{\theta}_{ij}) \big/ \hat{\sigma},$$

where the maximum operator *MAX* is taken over all possible points in the lattice

$$\{(i, j),\ i = 1,...,K,\ \ j = 1,...,N\}.$$

Both test statistics are then used by authors to test the hypothesis $H_0 : \theta_{ij} = 0 \;\; for \;\; all \;\; i, j \geq 1$ versus $H_1 : \theta_{ij} > 0 \;\; for \;\; some \;\; i, j \geq 1$. Note that these hypotheses are constructed for the case when the combination of the two drugs is not expected to be less effective than either component drug alone, *i.e.*, all $\theta_{ij} \geq 0$.

Using the fact that $Z_{ij} = \sqrt{n} \cdot (\bar{y}_{ij} - \mu_{ij}) / \sigma$ are distributed as standard normal random variables and the fact that given $\bar{y}_{i0}, \bar{y}_{0j}, \;\; for \;\; all \;\; i = 1,...,K, \;\; j = 1,...,N$, the variables $\hat{\theta}_{ij}$ are statistically independent and independent of $\hat{\sigma}$, the authors obtain the exact distribution functions of the test statistics T_A and T_{MAX}. The distribution function of AVE test statistic can be written as:

$$P\left(T_A \leq c \middle| \hat{\delta}, \hat{\theta}\right) = 1 - \int_0^{\infty} \int_{R^{K+N}} \Phi(h(u, v, w; \hat{\delta}, \hat{\theta})) \cdot \prod_{i=1}^{K} \varphi(u_i) \prod_{j=1}^{N} \varphi(v_j) du_1 ... du_K dv_1 ... dv_N dQ(w), \qquad (4.3)$$

where $\hat{\delta}$ and $\hat{\theta}$ are vectors with components δ_{ij}, θ_{ij} respectively, for $i = 1,...,K, \;\; j = 1,...,N$; $\Phi(\cdot)$ is the distribution function of the standard normal random variable; φ is its density function; $Q(\cdot)$ is the distribution function of $\hat{\sigma} / \sigma$, and R^{K+N} is the $(K+N)-$dimensional Euclidean space; the function

$$h(u, v, w; \hat{\delta}, \hat{\theta}) = -\sqrt{nKN}\left(cw - \frac{\theta_{AVE}}{\sigma}\right) - \frac{\sum_{i=1}^{K} \sum_{j=1}^{N} \left[u_i + v_j - \sqrt{n}\left|\delta_{ij}\right| + \left|u_i - v_j + \sqrt{n}\delta_{ij}\right|\right]}{2\sqrt{KN}},$$

where

$$\theta_{AVE} = \sum_{i=1}^{K} \sum_{j=1}^{N} \theta_{ij} \Big/ (K \cdot N), \;\; \delta_{ij} = (\mu_{i0} - \mu_{0j}) / \sigma.$$

Note that the distribution function of T_A depends on the parameters θ_{ij} only through θ_{AVE}.

The distribution function of the MAX test statistic can be written as

$$P\left(T_{MAX} \leq c \middle| \hat{\delta}, \hat{\theta}\right) = \int_0^{\infty} \int_{R^{K+N}} \prod_{i=1}^{K} \prod_{j=1}^{N} \Phi(g(u_i, v_j, w; \delta_{ij}, \theta_{ij})) \cdot \prod_{i=1}^{K} \varphi(u_i) \prod_{j=1}^{N} \varphi(v_j) du_1 ... du_K dv_1 ... dv_N dQ(w),$$

where $g(u_i, v_j, w; \delta_{ij}, \theta_{ij}) = \sqrt{n}\left(cw - \frac{\theta_{ij}}{\sigma}\right) + \frac{u_i + v_j - \sqrt{n}\left|\delta_{ij}\right| + \left|u_i - v_j + \sqrt{n}\delta_{ij}\right|}{2}$.

The above distribution functions involve the nuisance parameters δ_{ij}, so that the test statistics are not readily usable unless these parameters are given.

Next, by taking supremum over all $\delta_{ij} \in (-\infty, \infty)$ of the power functions of T_A and T_{MAX}, evaluated at $\widehat{\theta} = 0$, authors obtain the critical values for specified α by solving the following equations for c:

$$\alpha = \int_0^\infty \int_{-\infty}^\infty \Phi\left[-\sqrt{nKN}cw - \sqrt{a}u\right] \cdot \varphi(u) du dQ(w)$$

for the AVE test statistic, where $a = \max(K, N)$, and the equation

$$\alpha = 1 - \int_0^\infty E\left\{\left[\Phi(\sqrt{n}cw + Z)\right]^2\right\} \cdot \left\{E\left[\Phi(\sqrt{n}cw + Z)\right]\right\}^4 dQ(w)$$

for the MAX test, where E denotes the expectation and $Z \sim N(0,1)$. When variance is known, the authors provide the critical values of both standardized statistics, $\sqrt{n}T_A$ and $\sqrt{n}T_{MAX}$, for all rectangular designs with $K \le 5$ and $N \le 5$ for significance levels of $\alpha = 0.1,\ 0.05,\ 0.01$. Table 5 provides some of these values corresponding to the AVE test for $\alpha = 0.05$.

Table 5. α-level critical values of $\sqrt{n} \cdot T_A$, provided by Hung *et al.* (1993)

		Significance Level		
K	N	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$
1	1	1.81	2.33	3.29
1	2	1.57	2.01	2.85
1	3	1.48	1.90	2.69
2	2	1.11	1.42	2.01
2	3	1.05	1.34	1.90
3	3	0.85	1.10	1.55

The CPMP/EWP (2002) guidelines, formulated by the European Agency for the Evaluation of Medicinal Products state that there are two goals of employing multiple-dose factorial designs. These goals include "(i) to provide confirmatory evidence that a combination is

more effective than each of its components and (ii) to identify an effective and safe range of dose combinations for recommended use." Hellmich and Lehmacher (2005) consider the problem of identifying the minimum effective dose (MED) and do not address goal (i) of the CPMP/EWP (2002) guidelines.

Hellmich and Lehmacher (2005) introduce a definition of a minimum effective dose (MED) under the assumption that the response matrix has an isotonic structure. The doses of drug A and drug B are coded as $i,\ i = 0,...,K$ and $j,\ j = 0,...,N$ and the definition of an MED is formulated as $MED_{\prec} = \min_{\prec}\{(i,j) : \mu_{ij} > \mu_{00}\}$. The problem of identifying the $MED_{\prec}$ involves the following hypotheses

$$H_{0,ab} : \mu_{00} = \mu_{a'b'}\ \forall (a',b') \prec (a,b) \text{ versus}$$

$$H_{1,ab} : \mu_{00} \le \mu_{a'b'}\ \forall (a',b') \prec (a,b) \text{ and } \mu_{00} < \mu_{ab}.$$

The estimates for $MED_{\prec}$ are given by the smallest grid points (a,b) for which $H_{0,ab}$ is rejected. The authors construct the closure of family of the above hypotheses, which allows applying the Marcus, Peritz and Gabriel principle (Marcus *et al.* (1976)) to test each hypothesis from the closure at a fixed level α. Hellmich and Lehmacher consider the likelihood ratio and multiple contrast tests for testing each individual hypothesis. They also provide a tree-graph for the 2×2 design, illustrating the order in which the hypotheses should be tested.

In the remainder of this Chapter, we provide a step-down closed testing procedure for identifying all minimum efficacious doses (MeD's). By the efficacious combination we mean superiority of the combination, *i.e.*, the combination with positive expected gain. Our development begins with the simplest 2×2 case in order to present the proposed procedure in detail. We then consider the 2×3 case, which in turn introduces new complexities of the proposed procedure. The method for constructing the procedure for a general $K \times N$ case is very complex and does not have any practical value, so we provide the algorithm for constructing the procedure for any specified K and N and illustrate the algorithm for the 3×3 case.

4.2 CRITICAL VALUES FOR NON-RECTANGULAR DESIGNS

Before our procedure is discussed in detail we discuss non-rectangular designs and the methods for obtaining the critical values corresponding to these designs.

Hung *et al.* (1993) obtain the critical values of their AVE and MAX statistics for complete or rectangular designs. By a "complete" or "rectangular" design we mean the design, in which all possible combinations of two drugs are considered, *i.e.*, the combinations $G \equiv \{(i, j),\ i = 1,...,K,\ \ j = 1,...,N\}$. Our proposed procedure for identifying the MeD's which involves intersections of null hypotheses results in some of these hypotheses that do not correspond to the rectangular designs. So, in order to apply our procedure together with rectangular design we have to use some non-rectangular designs of the following form.

Let $k' \le K$, $n'_{(k')} \le n'_{(k'-1)} \le ... \le n'_{(1)} \le N$ with $n'_{(k')} < n'_{(1)}$. Then the non-rectangular design is given by

$$I \equiv \left\{(i, j), j = 1,...,n'_{(i)}, i = 1,...,k'\right\}. \tag{4.4}$$

The non-rectangular design I is obtained by the excluding some specified high dose combinations from the original grid of points G. Note that if in definition of I we do not require $n'_{(k')} < n'_{(1)}$, then $I \equiv G$.

The cardinality of G is KN and the cardinality of the lattice corresponding to non-rectangular design I is $n' = n'(k') = \sum_{i=1}^{k'} n'_{(i)}$, where numbers k', $n'_{(i)}$ depend on the actual values of K and N and are to be specified for our testing procedure. Now, let us just mention that the simplest case of non-rectangular design corresponds to the case when a single combination is omitted from the original set of points. There are KN possible such non-rectangular designs, but we are interested only in a single one of them, where the highest drug combination (K, N) is omitted from the original grid G. For the 2×2 case, it turns out that this form of non-rectangular design is the only one which we need to consider, and is given by $I_{2\times 2} \equiv \left\{\{(i, j = j_{(i)}), i = 1,...,k', j_{(i)} = 1,...,n'_{(i)}\},\ \ for\ k' = 2,\ n'_{(1)} = 2,\ n'_{(2)} = 1\right\}$, *i.e.*, the set of drug combinations $\{(1,1),(1,2),(2,1)\}$.

In order to obtain the critical values for the specified non-rectangular design I, we need to generalize Hung's *et al.* (1993) method for obtaining the critical value for the AVE test. Assume that I is given, so that all of k', $n'_{(i)}$, $i = 1,...,k'$ (with $n'_{(1)} \geq 2$), n' (the total number of drug combinations under consideration) and $P = k' + n'_{(1)}$ (the total number of single drug doses) are specified. Then the following theorem holds (the detailed proof is given in Appendix D).

<u>Theorem 4.2</u>. Let $k' \leq K$, $n'_{(k')} \leq n'_{(k'-1)} \leq ... \leq n'_{(1)}$ with $n'_{(k')} < n'_{(1)}$ and $n' = \sum_{i=1}^{k'} n'_{(i)}$ and define $G(t) = \int_0^{\infty} \int_{-\infty}^{\infty} \Phi\left[-\sqrt{n \cdot n'}cw - tx\right] \varphi(x) dx dQ(w)$. Then for non-rectangular design I of the form (4.4), the α-level critical value of the AVE test statistic

$$T_A = \frac{\sum_{(i.j) \in I} \hat{\theta}_{ij}}{n' \cdot \hat{\sigma}}$$

is given as the solution c to the equation $G(t_{\max}) = \alpha$, where $t_{\max} = \max \sqrt{\frac{\sum_{s=1}^{P} m_s^2}{n'}}$ with m_s, $s = 1,...,P$ non-negative integers subject to constraints:

$$m_s \leq n'_{(s)}, \quad s = 1,...,k';$$
$$m_s \leq \left|\left\{(i, s-k') : (i, s-k') \in I\right\}\right|, \quad s = k'+1,...,P$$

and $\sum_{s=1}^{P} m_s = n'$, where $|A|$ denotes the cardinality of a set A.

In order to use Theorem 4.2, first one needs to solve the optimization problem of maximizing the value of $t = \sqrt{\frac{\sum_{s=1}^{P} m_s^2}{n'}}$ subject to specific constraints on the values of m_s, $s = 1,...,P$. The solution to this problem is given in Appendix E. Next, the equation $G(t_{\max}) = \alpha$ must be solved for c, for the specified value of α; this can be done using *MathCAD, Maple* or another mathematical software package. In the case of known variance, the double integral in the

expression for $G(t)$ reduces to $G(t) = \int_{-\infty}^{\infty} \Phi[-\sqrt{n \cdot n'}c - tx]\varphi(x)dx$ and can be used to obtain the critical values.

Example 4.1. For the non-rectangular design of the form, $I_0 = G \setminus \{(K,N)\} \equiv \{\{(i,j), i = 1,...,K, j = 1,...,N\} \setminus \{(K,N)\}\}$, the maximum value of $t = \sqrt{\frac{\sum_{s=1}^{P} m_s^2}{n'}}$ subject to above constraints equals $t_{\max} = \sqrt{\frac{ab^2 - 2b + 1}{n'}}$, where $a = \min\{K,\ N\}$ and $b = \max\{K,\ N\}$ and $n' = KN - 1$. The proof of this fact is given in Appendix E. Then, the significance level of the AVE test is given by

$$Significance\ \ Level \equiv \int_0^{\infty}\int_{-\infty}^{\infty} \Phi\left[-\sqrt{n(KN-1)}cw - \sqrt{\frac{ab^2 - 2b + 1}{KN-1}} \cdot x\right]\varphi(x)dxdQ(w).$$

So the critical values of desired level of significance can be obtained by solving the following equation for c:

$$\alpha = \int_0^{\infty}\int_{-\infty}^{\infty} \Phi\left[-\sqrt{n(KN-1)}cw - \sqrt{\frac{ab^2 - 2b + 1}{KN-1}} \cdot x\right]\varphi(x)dxdQ(w), \tag{4.5a}$$

which simplifies to

$$\alpha = \int_{-\infty}^{\infty} \Phi(-\sqrt{n(KN-1)} \cdot c - \sqrt{\frac{ab^2 - 2b + 1}{KN-1}} \cdot u) \cdot \varphi(u)du \tag{4.5b}$$

when variance is known.

This is the only form of the non-rectangular design which is used by us in the 2×2 case, and the corresponding critical values for this case are given in the Section 4.3. All other higher dimensional designs require numerical optimization to compute the critical values.

4.3 2 BY 2 CASE

Next, we are going to construct a stepwise procedure for identifying the minimum efficacious doses. To illustrate the approach and its complexity we first discuss its details for the 2×2 case.

Our procedure involves the test statistics proposed by Hung *et al.* (1993). First, for a given design all possible MeD-sets must be identified. Next, we identify the corresponding closed family of hypotheses. Our procedure uses a graph-theoretic testing order for the hypotheses so we can employ the Marcus, Peritz and Gabriel (1976) closed testing principle. This principle along with a prescribed step-down testing allows us to estimate the MeD-set.

In order to test an individual hypothesis for each active compound, the expected gain from combining the drugs at the given levels is defined. Next, the appropriate estimator of the expected gain is constructed in order to test the hypotheses of the expected gains being zero versus at least one positive gain. The pattern of the rejected null hypotheses identifies all the combinations, *i.e.*, the MeD's, such that decreasing at least one component's dose leads to no-gain compound.

Although, we consider a general $K \times N$ combination drug set up, in most biopharmaceutical studies, the number of drug dose levels is not typically bigger than four or five, because of the high cost of the study and the large number of participants needed if the number of experimental groups is large. We begin with the simplest 2×2 case.

Denote the matrix of unknown expected gains in the case of 2×2 design by $\Theta = \begin{pmatrix} \theta_{11} & \theta_{12} \\ \theta_{21} & \theta_{22} \end{pmatrix}$. First, the collection of all possible MeD-sets for the 2×2 case should be identified. This collection is given by (3.1) and consists of the sets represented by the following matrices:

$$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \text{ and } \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

Next, we construct a hypothesis family $\mathrm{H} = \left\{H_0^{(4)}, H_0^{(3)}, H_0^{(2')}, H_0^{(2'')}, H_0^{(1)}\right\}$ where

$H_0^{(4)} : \theta_{11} = \theta_{12} = \theta_{21} = \theta_{22} = 0$,

$H_0^{(3)} : \theta_{11} = \theta_{12} = \theta_{21} = 0$,

$H_0^{(2')} : \theta_{11} = \theta_{12} = 0$, $H_0^{(2'')} : \theta_{11} = \theta_{21} = 0$

$H_0^{(1)} : \theta_{11} = 0$

with the corresponding alternative hypotheses given by

$H_1^{(4)} : \theta_{11} \geq 0,\ \theta_{12} \geq 0,\ \theta_{21} \geq 0,\ \theta_{22} \geq 0$ *with at least one* ">"

$H_1^{(3)}: \theta_{11} \geq 0,\ \theta_{12} \geq 0,\ \theta_{21} \geq 0,\ \textit{with at least one}$ ">"

$H_1^{(2')}: \theta_{11} \geq 0,\ \theta_{12} \geq 0\ \textit{with at least one}$ ">"

$H_1^{(2'')}: \theta_{11} \geq 0,\ \theta_{21} \geq 0\ \textit{with at least one}$ ">"

$H_1^{(1)}: \theta_{11} > 0$.

Since, $H_0^{(4)} \subset H_0^{(3)} \subset H_0^{(2')} \subset H_0^{(1)}$, $H_0^{(4)} \subset H_0^{(3)} \subset H_0^{(2'')} \subset H_0^{(1)}$ and $H_0^{(2')} \bigcap H_0^{(2'')} \equiv H_0^{(3)}$ then the intersection of any two hypotheses $H_0^{(s)} \in \mathrm{H},\ H_0^{(t)} \in \mathrm{H},\ s\ \textit{is not identical to}\ t$, is given by

$$H_0^{(s)} \bigcap H_0^{(t)} = \begin{cases} H_0^{(s)},\ \textit{if}\ s > t \\ H_0^{(t)},\ \textit{if}\ s < t \\ H_0^{(3)},\ \textit{otherwise} \end{cases}$$

which belongs to the family; if $s\ \textit{is identical to}\ t,\ H_0^{(s)} \bigcap H_0^{(t)} \equiv H_0^{(s)} \in \mathrm{H}$. Hence, the family of hypotheses H is closed under intersection. Now the Marcus, Peritz and Gabriel method for constructing a procedure that strongly controls the Familywise Error (Marcus *et al.* (1976)) can be applied under the rule that a hypothesis is rejected at level α if and only if all hypotheses implying it, including itself, are significant at level α.

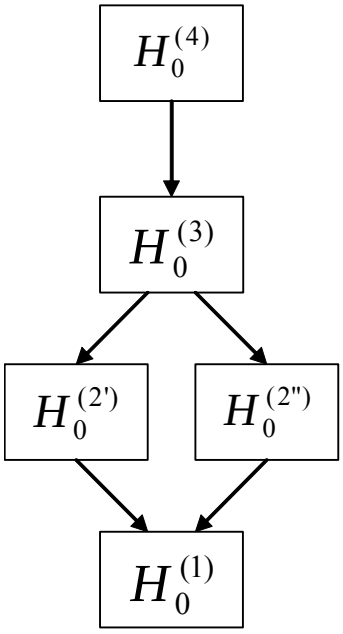

Figure 4. The "implication" tree for the 2×2 design

The "implication of rejection" tree for the 2×2 design is given in Figure 4. A similar tree for the 2×2 design was first proposed by Hellmich *et al.* (2005) but for a different problem set-up. The tree given in Figure 4 has five nodes and four levels, which we call "hypothesis levels", labeled from bottom to top. The fourth, third and first hypothesis levels consist of a single hypothesis, but the second hypothesis level has two hypotheses. This tree defines the order in which the hypotheses are to be tested (each hypothesis is tested at level α). First, we test $H_0^{(4)}$, if $H_0^{(4)}$ is accepted then we stop and conclude that there are no MeD's, otherwise we test $H_0^{(3)}$. If $H_0^{(3)}$ is accepted then we stop and conclude that the MeD-matrix is estimated as

$$\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

If $H_0^{(3)}$ is rejected then we test both hypotheses $H_0^{(2')}$ and $H_0^{(2'')}$. If both hypotheses are rejected we test $H_0^{(1)}$. If $H_0^{(2')}$ is accepted and $H_0^{(2'')}$ is rejected then we conclude that the MeD-matrix is estimated as

$$\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}.$$

In the case of rejecting $H_0^{(2')}$ and accepting $H_0^{(2'')}$, the MeD-matrix is estimated as

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

Finally, in the case of testing $H_0^{(1)}$ if it is accepted than we conclude that there are two MeD's, *i.e.*, the MeD-matrix is estimated as

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

and if $H_0^{(1)}$ is rejected then it is estimated as

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

The above conclusions are summarized in Table 6. Note that the case when both $H_0^{(2')}$ and $H_0^{(2'')}$ are accepted given that $H_0^{(3)}$ was rejected, does not provide any conclusive result in

terms of the estimated MeD's. In such cases we say that the procedure results in Type A ambiguity. This situation is considered in more detail in Section 4.6.

Each of the hypotheses from the family H can be tested using Hung's *et al.* (1993) AVE test statistic, given by (4.2). The additional critical values are needed only for non-rectangular design of testing $H_0^{(3)}: \theta_{11} = \theta_{12} = \theta_{21} = 0$, so the results (4.5a) and (4.5b) can be used. Applying (4.5a), we obtain the critical values by solving the equation below for c for the specified level of significance:

$$\alpha = \int_0^{\infty} \int_{-\infty}^{\infty} \Phi\left[-\sqrt{3n}cw - \sqrt{5/3} \cdot u\right] \cdot \varphi(u) du dQ(w) .$$

Table 6. "Decision" sequences with the corresponding estimated MeD-sets for the 2×2 case (Notation: ACC - "Hypothesis is accepted", REJ – "Hypothesis is rejected", NT -"Hypothesis is not tested.")

$H_0^{(4)}$	$H_0^{(3)}$	$H_0^{(2')}$	$H_0^{(2'')}$	$H_0^{(1)}$	Estimated $MED-set$
ACC	NT	NT	NT	NT	Empty
REJ	ACC	NT	NT	NT	{(2,2)}
REJ	REJ	ACC	ACC	NT	Type A
REJ	REJ	ACC	REJ	NT	{(2,1)}
REJ	REJ	REJ	ACC	NT	{(1,2)}
REJ	REJ	REJ	REJ	ACC	{(2,1), (1,2)}
REJ	REJ	REJ	REJ	REJ	{(1,1)}

In the case of known variance, from (4.5b) it follows that the critical values can be obtained from $\alpha = \int_{-\infty}^{\infty} \Phi(-\sqrt{3n} \cdot c - \sqrt{5/3} \cdot u) \cdot \varphi(u) du$.

For the standardized AVE test, the critical values corresponding to $\alpha = 0.1,\ 0.05,\ 0.01$ obtained from the last equation are, respectively, $c_{0.1} = 1.21,\ c_{0.05} = 1.55,\ c_{0.01} = 2.19$.

Example 4.3. Hung *et al.* (1993) provide data from a factorial clinical trial, which is conducted to evaluate the effectiveness of the combination of two antihypertensive drugs, A and B. The

subset of observed data of diastolic blood pressure mean reductions (in mm Hg, rounded to the nearest integer) is given below.

	Drug B		
Drug A	0	1	2
0	0	4	5
1	5	9	7
2	5	6	6

The sample size of each group is 25. The observed blood pressure reductions appear to be approximately normally distributed. The pooled estimate for the true variance is 42.

To illustrate our procedure we consider the above data of diastolic blood pressure mean reductions for two active dose levels of each drug. The matrix of estimated gains is $\hat{\Theta} = \begin{pmatrix} 4 & 2 \\ 1 & 1 \end{pmatrix}$.

The AVE test statistics corresponding to $H_0^{(4)}$, $H_0^{(3)}$, $H_0^{(2')}$, $H_0^{(2'')}$ and $H_0^{(1)}$ and the critical values are

$$T_A^{(4)} = \frac{4+2+1+1}{2 \cdot 2 \cdot \sqrt{42}} = 0.31,\ c_{0.05}^{(4)} = 1.42/\sqrt{25} = 0.28,$$

$$T_A^{(3)} = \frac{4+2+1}{3 \cdot \sqrt{42}} = 0.36,\ c_{0.05}^{(3)} = 1.55/\sqrt{25} = 0.31.$$

The average of three observed gains is taken to calculate $T_A^{(3)}$ so the expression in the denominator, $K \cdot N$, is substituted by 3.

$$T_A^{(2')} = \frac{4+2}{2 \cdot \sqrt{42}} = 0.46,\ c_{0.05}^{(2')} = c_{0.05}^{(2'')} = 2.01/\sqrt{25} = 0.40,$$

$$T_A^{(2'')} = \frac{4+1}{2 \cdot \sqrt{42}} = 0.39,$$

$$T_A^{(1)} = \frac{4}{1 \cdot \sqrt{42}} = 0.62,\ c_{0.05}^{(1)} = 2.33/\sqrt{25} = 0.47.$$

Since $T_A^{(4)} > c_{0.05}^{(4)}$ we reject $H_0^{(4)} : \theta_{11} = \theta_{12} = \theta_{21} = \theta_{22} = 0$. Then, we test $H_0^{(3)} : \theta_{11} = \theta_{12} = \theta_{21} = 0$ and since $T_A^{(3)} > c_{0.05}^{(3)}$, we reject it. Next, we test and reject $H_0^{(2')} : \theta_{11} = \theta_{12} = 0$ and test and accept $H_0^{(2'')} : \theta_{11} = \theta_{21} = 0$. So we stop and conclude that the estimated MeD-set is given by the set $\{(1,2)\}$.

4.4 2 BY 3 CASE

We next consider the 2×3 design, which illustrates some additional complexities not apparent in the 2×2 case. Before we introduce the hypothesis family for this case, all nine possible MeD-sets must be identified. First, there is an empty set, six different sets of a single element $\{(1,1)\}$, $\{(1,2)\}$, $\{(1,3)\}$, $\{(2,1)\}$, $\{(2,2)\}$, $\{(2,3)\}$ and there are three different sets of two elements $\{(1,3),\ (2,2)\}$, $\{(2,1),\ (1,2)\}$ and $\{(2,2),\ (1,3)\}$.

For the matrix of population gains $\Theta = \begin{pmatrix} \theta_{11} & \theta_{12} & \theta_{13} \\ \theta_{21} & \theta_{22} & \theta_{23} \end{pmatrix}$ we consider the set of null hypotheses (the corresponding one-sided alternative hypotheses are parallel to the 2×2 case)

$H_0^{(6)} : \theta_{11} = \theta_{12} = \theta_{13} = \theta_{21} = \theta_{22} = \theta_{23} = 0$

$H_0^{(5)} : \theta_{11} = \theta_{12} = \theta_{13} = \theta_{21} = \theta_{22} = 0$

$H_0^{(4')} : \theta_{11} = \theta_{12} = \theta_{13} = \theta_{21} = 0$

$H_0^{(4'')} : \theta_{11} = \theta_{12} = \theta_{21} = \theta_{22} = 0$

$H_0^{(3')} : \theta_{11} = \theta_{12} = \theta_{13} = 0$

$H_0^{(3'')} : \theta_{11} = \theta_{12} = \theta_{21} = 0$

$H_0^{(2')} : \theta_{11} = \theta_{12} = 0$

$H_0^{(2'')} : \theta_{11} = \theta_{21} = 0$

$H_0^{(1)} : \theta_{11} = 0$.

The testing order is represented by the "implication" tree of Figure 5.

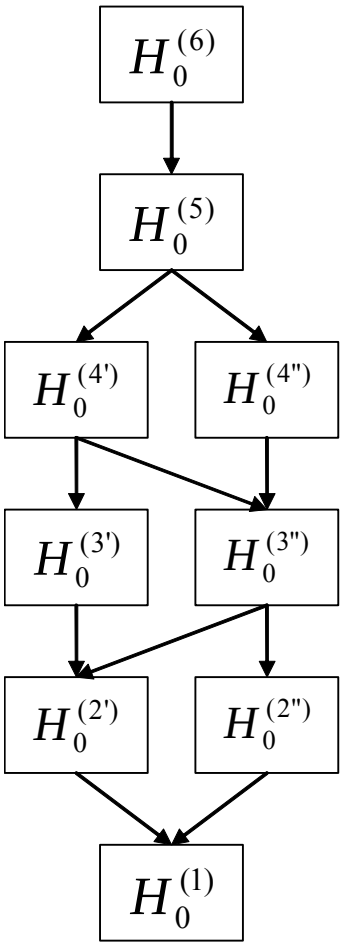

Figure 5. The "implication" tree for the 2×3 design

We use this tree to show that the family of hypotheses $\mathrm{H} = \left\{H_0^{(6)}, H_0^{(5)}, H_0^{(4')}, H_0^{(4'')}, H_0^{(3')}, H_0^{(3'')}, H_0^{(2')}, H_0^{(2'')}, H_0^{(1)}\right\}$ is closed under intersection. Indeed, the intersection of any number of hypotheses connected with implication arrows, $H_0^{(s_1)}$, $H_0^{(s_2)}, \ldots, H_0^{(s_h)}$ for some h, $1 < h \le 6$ is $\bigcap_{\substack{s_1<s_2<\ldots<s_h \\ 1\le i\le h}} H_0^{(s_i)} = H_0^{(s_h)}$. The intersection of any hypotheses of the same level is the hypothesis of the higher level, indeed, $H_0^{(4')} \bigcap H_0^{(4'')} \equiv H_0^{(5)}$, $H_0^{(3')} \bigcap H_0^{(3'')} \equiv H_0^{(4')}$ and $H_0^{(2')} \bigcap H_0^{(2'')} \equiv H_0^{(3'')}$. So the only hypotheses whose intersection is left to be considered are the hypotheses which do not imply one another and are of the different levels. There are two pairs of such hypotheses, $\left\{H_0^{(4'')}, H_0^{(3')}\right\}$ and $\left\{H_0^{(3')}, H_0^{(2'')}\right\}$. Since $H_0^{(4'')} \bigcap H_0^{(3')} \equiv H_0^{(5)}$ and $H_0^{(3')} \bigcap H_0^{(2'')} \equiv H_0^{(4')}$ both intersection-hypotheses belong to the family, which completes the proof that H is closed under intersection.

Table 7 illustrates that our procedure produces the estimates of all possible MeD-sets. Unfortunately, together with these conclusive results, the procedure can result in other "decision" sequences of rejected and accepted hypotheses, which lead to ambiguities in terms of the estimated MeD-set. These situations are of two kinds, Type A and Type B. One method that can be used to deal with Type B ambiguity is presented in Section 4.6.

In order to test the above hypotheses using the AVE test statistic (4.2), we need to calculate the additional critical values for non-rectangular designs in order to test $H_0^{(5)}$, $H_0^{(4')}$ and $H_0^{(3'')}$. The critical values for testing $H_0^{(3'')}$ based on the AVE test were discussed in the 2×2 case. For testing $H_0^{(5)}$ the results (4.5a) and (4.5b) provide the following equations:

$\alpha = \int_0^{\infty} \int_{-\infty}^{\infty} \Phi\left[-\sqrt{5n}cw - \sqrt{13/5} \cdot u\right] \cdot \varphi(u) du dQ(w)$ (if the variance is not known) and

$\alpha = \int_{-\infty}^{\infty} \Phi(-\sqrt{5n} \cdot c - \sqrt{13/5} \cdot u) \cdot \varphi(u) du$ if the variance is known.

Table 7. "Decision" sequences with the corresponding estimated MeD-sets for the 2×3 case (Notation: ACC - "Hypothesis is accepted", REJ – "Hypothesis is rejected", NT -"Hypothesis is not tested")

$H_0^{(6)}$	$H_0^{(5)}$	$H_0^{(4')}$	$H_0^{(4'')}$	$H_0^{(3')}$	$H_0^{(3'')}$	$H_0^{(2')}$	$H_0^{(2'')}$	$H_0^{(1)}$	Estimated MeD-set
ACC	NT	NT	NT	NT	NT	NT	NT	NT	Empty
REJ	ACC	NT	NT	NT	NT	NT	NT	NT	{(2,3)}
REJ	REJ	ACC	ACC	NT	NT	NT	NT	NT	Type A
REJ	REJ	ACC	REJ	NT	NT	NT	NT	NT	{(2,2)}
REJ	REJ	REJ	ACC	REJ	NT	NT	NT	NT	{(1,3)}
REJ	REJ	REJ	ACC	ACC	NT	NT	NT	NT	Type B
REJ	REJ	REJ	REJ	ACC	ACC	NT	NT	NT	Type A
REJ	REJ	REJ	REJ	ACC	REJ	NT	REJ	NT	{(2,1)}
REJ	REJ	REJ	REJ	ACC	REJ	NT	ACC	NT	Type B
REJ	REJ	REJ	REJ	REJ	ACC	NT	NT	NT	{(2,2), (1,3)}
REJ	REJ	REJ	REJ	REJ	REJ	ACC	ACC	NT	Type A
REJ	REJ	REJ	REJ	REJ	REJ	REJ	ACC	NT	{(1,2)}
REJ	REJ	REJ	REJ	REJ	REJ	ACC	REJ	NT	{(2,1), (1,3)}
REJ	REJ	REJ	REJ	REJ	REJ	REJ	REJ	ACC	{(2,1), (1,2)}
REJ	REJ	REJ	REJ	REJ	REJ	REJ	REJ	REJ	{(1,1)}

For example, for the standardized AVE test the critical values corresponding to $\alpha = 0.1,\ 0.05,\ 0.01$ obtained from the last equation are, respectively, $c_{0.1} = 1.09$, $c_{0.05} = 1.40$ and $c_{0.01} = 1.97$. For testing $H_0^{(4')}$ the α-level critical value of the AVE test can be obtained as the solution to the following equation for the specified level of significance α:

$\alpha = \int_0^{\infty} \int_{-\infty}^{\infty} \Phi\left[-\sqrt{4n}cw - \sqrt{10/4}x\right] \varphi(x)dxdQ(w)$ (if variance is unknown) and from

$\alpha = \int_{-\infty}^{\infty} \Phi(-\sqrt{4n} \cdot c - \sqrt{10/4} \cdot u) \cdot \varphi(u)du$ (if variance is known). For example, in the case of known variance the critical values are given by $c_{0.1} = 1.20$, $c_{0.05} = 1.54$ and $c_{0.01} = 2.18$ for the standardized AVE test.

The details on obtaining the above integrals are considered in Appendix E.

4.5 3 BY 3 CASE

In the cases of 2×2 and 2×3 designs the number of different MeD-sets and the number of elements in each set are not relatively large, so it is quite easy to construct a sequence of hypotheses which would produce the estimates of all possible MeD-sets configurations and be closed at the same time. When the number of drug dose levels increases, the structure of each MeD-set becomes more complicated. We consider the case of the 3×3 design and illustrate a general approach for constructing such a procedure. The same techniques can be used for analysis of a combination drug in the 2×4, 3×4, 4×4 trials and trials of higher dimensions.

First, all possible MeD-sets must be identified and the corresponding hypotheses constructed. Then, the proper testing order must be set up. Next, for all non-rectangular types of design the additional critical values need to be calculated. Finally, the step-down testing procedure can be applied using the MPG principle or any modification of it, which guarantees strong control of the FWE.

Let us look at all possible MeD-sets for the 3×3 design. The cardinality, p, of the MeD-set can be any number such that $1 \le p \le \min\{3,\ 3\} = 3$. We use the recursive formula given by (3.2) to check that all possible MeD-sets are identified for each p.

There are nine different MeD-sets with $p=1$, they are $\{(1,1)\}$, $\{(1,2)\}$, $\{(1,3)\}$, $\{(2,1)\}$, $\{(2,2)\}$, $\{(2,3)\}$, $\{(3,1)\}$, $\{(3,2)\}$ and $\{(3,3)\}$. For $p=2$, the formula gives us

$$q(2,3,3)=(3\cdot(3-1)\cdot 3\cdot(3-1))/4=9$$

MeD-sets. We identify them as $\{(1,2),\ (2,1)\}$, $\{(1,2),\ (3,1)\}$, $\{(1,3),\ (2,1)\}$, $\{(1,3),\ (2,2)\}$, $\{(1,3),\ (3,1)\}$, $\{(1,3),\ (3,2)\}$, $\{(2,2),\ (3,1)\}$, $\{(2,3),\ (3,1)\}$ and $\{(2,3),\ (3,2)\}$.

Finally, there is a single set $\{(1,3\},\ (2,2),\ (3,1)\}$ of cardinality $p=3$. Thus, in the 3×3 case there are a total of 19 possible non-empty MeD-sets.

Now, let us construct the family of the null hypotheses. Each corresponding alternative hypothesis states that all gains considered in the null hypothesis are non-negative and at least one of them is strictly positive. The null hypothesis of the highest level includes all nine gains. By excluding the highest combination $(3,3)$, we form the hypothesis of the eighth level. The next highest combinations are $(2,3)$ and $(3,2)$, so we form two hypotheses of the seventh level by omitting one of these combinations. Then, we continue to exclude one highest dose at a time from each hypothesis, until we are left with the only one gain, corresponding to the null hypothesis of the first level $H_0^{(1)}:\theta_{11}=0$. The family of hypotheses constructed in such a way is presented below, where the superscript denotes the level from lowest (1) to largest (9).

$H_0^{(1)}:\theta_{11}=0$

$H_0^{(2')}:\theta_{11}=\theta_{12}=0$

$H_0^{(2'')}:\theta_{11}=\theta_{21}=0$

$H_0^{(3''')}:\theta_{11}=\theta_{12}=\theta_{13}=0$

$H_0^{(3'')}:\theta_{11}=\theta_{21}=\theta_{31}=0$

$H_0^{(3')}:\theta_{11}=\theta_{12}=\theta_{21}=0$

$H_0^{(4')}:\theta_{11}=\theta_{12}=\theta_{21}=\theta_{22}=0$

$H_0^{(4'')}:\theta_{11}=\theta_{12}=\theta_{21}=\theta_{31}=0$

$H_0^{(4''')}:\theta_{11}=\theta_{12}=\theta_{13}=\theta_{21}=0$

$H_0^{(5')}:\theta_{11}=\theta_{12}=\theta_{21}=\theta_{22}=\theta_{31}=0$

$H_0^{(5'')}:\theta_{11}=\theta_{12}=\theta_{13}=\theta_{21}=\theta_{31}=0$

$H_0^{(5''')}: \theta_{11} = \theta_{12} = \theta_{13} = \theta_{21} = \theta_{22} = 0$

$H_0^{(6')}: \theta_{11} = \theta_{12} = \theta_{21} = \theta_{22} = \theta_{31} = \theta_{32} = 0$

$H_0^{(6'')}: \theta_{11} = \theta_{12} = \theta_{13} = \theta_{21} = \theta_{22} = \theta_{32} = 0$

$H_0^{(6''')}: \theta_{11} = \theta_{12} = \theta_{13} = \theta_{21} = \theta_{22} = \theta_{23} = 0$

$H_0^{(7')}: \theta_{11} = \theta_{12} = \theta_{13} = \theta_{21} = \theta_{22} = \theta_{31} = \theta_{32} = 0$

$H_0^{(7'')}: \theta_{11} = \theta_{12} = \theta_{13} = \theta_{21} = \theta_{22} = \theta_{23} = \theta_{31} = 0$

$H_0^{(8)}: \theta_{11} = \theta_{12} = \theta_{13} = \theta_{21} = \theta_{22} = \theta_{23} = \theta_{31} = \theta_{32} = 0$

$H_0^{(9)}: \theta_{11} = \theta_{12} = \theta_{13} = \theta_{21} = \theta_{22} = \theta_{23} = \theta_{31} = \theta_{32} = \theta_{33} = 0$.

Again, let H be the family of these null hypotheses. Then it can be shown that H is closed under intersection, so the MPG principle can be used to test the above sequence of hypotheses and identify the MeD-set. The order of testing scheme then can be represented by the "implication" tree given in Figure 6. There are 20 different MeD-sets (including the empty set) for the 3×3 design and Table 8 illustrates that our procedure produces the estimates of all possible population MeD-sets. As in the 2×2 and 2×3 cases, the procedure can result in other "decision" sequences of rejected and accepted hypotheses leading to ambiguities. Such sequences and some suggestions on how to deal with these situations are given in Section 4.6.

In order to test the above hypotheses using the AVE test statistic (4.2), we need to calculate the additional critical values for non-rectangular designs in order to test $H_0^{(8)}$, $H_0^{(7')}$, $H_0^{(7'')}$, $H_0^{(6'')}$, $H_0^{(5')}$, $H_0^{(5'')}$, $H_0^{(5''')}$, $H_0^{(4'')}$, $H_0^{(4''')}$ and $H_0^{(3')}$. Using the results of (4.5a) and (4.5b), we can obtain the critical values for testing $H_0^{(8)}$, $H_0^{(5''')}$ and $H_0^{(3')}$. Theorem 4.2 of Section 4.2 should be used in order to test $H_0^{(7')}$, $H_0^{(7'')}$, $H_0^{(6'')}$, $H_0^{(5')}$, $H_0^{(5'')}$, $H_0^{(4'')}$ and $H_0^{(4''')}$. After all the additional critical values are obtained, the step-down testing procedure can be applied using the MPG principle or any modification of it, which guarantees strong control of the FWE.

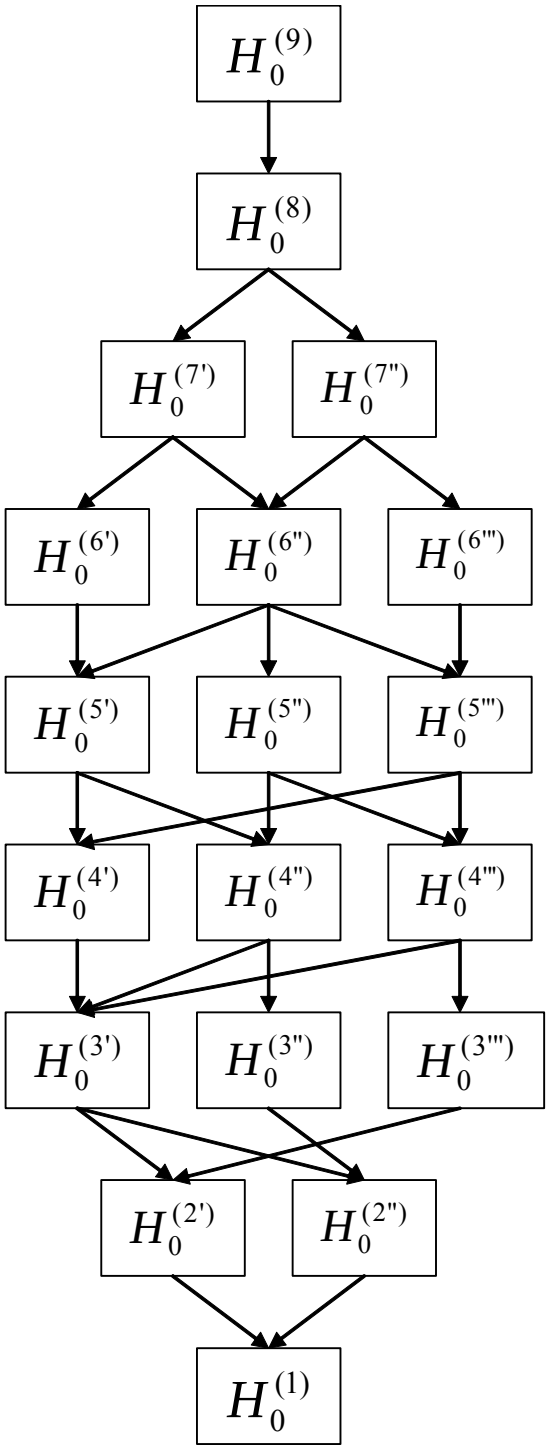

Figure 6. The "implication" tree for the 3×3 design

Table 8. "Decision" sequences with the corresponding estimated MeD-sets for the 3×3 case (Notation: ACC - "Hypothesis is accepted", REJ – "Hypothesis is rejected", NT -"Hypothesis is not tested", $\hat{M}$ - "Estimated MeD-set")

$H_0^{(9)}$	ACC	REJ	REJ	REJ	REJ	REJ	REJ	REJ	REJ	REJ
$H_0^{(8)}$	NT	REJ	REJ	REJ	REJ	REJ	REJ	REJ	REJ	ACC
$H_0^{(7')}$	NT	REJ	REJ	REJ	REJ	REJ	ACC	REJ	REJ	NT
$H_0^{(7'')}$	NT	REJ	REJ	REJ	REJ	REJ	REJ	REJ	ACC	NT
$H_0^{(6')}$	NT	REJ	REJ	ACC	REJ	REJ	NT	REJ	REJ	NT
$H_0^{(6'')}$	NT	REJ	REJ	REJ	REJ	REJ	NT	REJ	NT	NT
$H_0^{(6''')}$	NT	REJ	REJ	REJ	REJ	REJ	REJ	ACC	NT	NT
$H_0^{(5')}$	NT	REJ	REJ	NT	REJ	REJ	NT	REJ	NT	NT
$H_0^{(5'')}$	NT	REJ	REJ	REJ	REJ	ACC	NT	REJ	NT	NT
$H_0^{(5''')}$	NT	REJ	REJ	REJ	REJ	REJ	NT	NT	NT	NT
$H_0^{(4')}$	NT	REJ	REJ	NT	REJ	REJ	NT	NT	NT	NT
$H_0^{(4'')}$	NT	REJ	REJ	NT	REJ	NT	NT	REJ	NT	NT
$H_0^{(4''')}$	NT	REJ	REJ	REJ	REJ	NT	NT	NT	NT	NT
$H_0^{(3')}$	NT	REJ	REJ	NT	REJ	NT	NT	NT	NT	NT
$H_0^{(3'')}$	NT	REJ	ACC	NT	REJ	NT	NT	REJ	NT	NT
$H_0^{(3''')}$	NT	REJ	REJ	REJ	ACC	NT	NT	NT	NT	NT
$H_0^{(2')}$	NT	REJ	REJ	NT	NT	NT	NT	NT	NT	NT
$H_0^{(2'')}$	NT	REJ	NT	NT	REJ	NT	NT	NT	NT	NT
$H_0^{(1)}$	NT	REJ	NT	NT	NT	NT	NT	NT	NT	NT
$\hat{M}$	Empty	(1,1)	(1,2)	(1,3)	(2,1)	(2,2)	(2,3)	(3,1)	(3,2)	(3,3)

Table 8 continued

$H_0^{(9)}$	REJ	REJ	REJ	REJ	REJ	REJ	REJ	REJ	REJ	REJ
$H_0^{(8)}$	REJ	REJ	REJ	REJ	REJ	REJ	REJ	REJ	REJ	REJ
$H_0^{(7')}$	REJ	REJ	REJ	REJ	REJ	REJ	REJ	REJ	REJ	REJ
$H_0^{(7'')}$	REJ	REJ	REJ	REJ	REJ	REJ	REJ	REJ	REJ	REJ
$H_0^{(6')}$	REJ	REJ	REJ	REJ	REJ	REJ	REJ	REJ	REJ	REJ
$H_0^{(6'')}$	REJ	REJ	REJ	REJ	REJ	REJ	REJ	REJ	ACC	REJ
$H_0^{(6''')}$	REJ	REJ	REJ	REJ	REJ	REJ	REJ	REJ	REJ	REJ
$H_0^{(5')}$	REJ	REJ	REJ	REJ	REJ	ACC	REJ	REJ	NT	REJ
$H_0^{(5'')}$	REJ	REJ	REJ	REJ	REJ	REJ	REJ	REJ	NT	REJ
$H_0^{(5''')}$	REJ	REJ	REJ	REJ	REJ	REJ	REJ	ACC	NT	REJ
$H_0^{(4')}$	REJ	REJ	REJ	REJ	ACC	NT	REJ	NT	NT	REJ
$H_0^{(4'')}$	REJ	REJ	REJ	ACC	REJ	NT	REJ	REJ	NT	REJ
$H_0^{(4''')}$	REJ	REJ	REJ	REJ	REJ	REJ	ACC	NT	NT	REJ
$H_0^{(3')}$	REJ	REJ	REJ	NT	NT	NT	NT	NT	NT	ACC
$H_0^{(3'')}$	REJ	REJ	REJ	NT	REJ	NT	REJ	REJ	NT	REJ
$H_0^{(3''')}$	REJ	REJ	REJ	REJ	REJ	REJ	NT	NT	NT	REJ
$H_0^{(2')}$	REJ	REJ	ACC	NT	NT	NT	NT	NT	NT	NT
$H_0^{(2'')}$	REJ	ACC	REJ	NT	NT	NT	NT	NT	NT	NT
$H_0^{(1)}$	ACC	NT	NT	NT	NT	NT	NT	NT	NT	NT
$\hat{M}$	(1,2) (2,1)	(1,2) (3,1)	(1,3) (2,1)	(1,3) (2,2)	(1,3) (3,1)	(1,3) (3,2)	(2,2) (3,1)	(2,3) (3,1)	(2,3) (3,2)	(1,3) (2,2) (3,1)

4.6 ON AMBIGUITIES OF THE PROPOSED PROCEDURE

We have shown that our procedure provides estimators which include all possible MeD-sets in the 2×2, 2×3 and 3×3 cases, but we have also mentioned that there are some "decision" sequences of accepted and rejected hypotheses which lead to inconclusive or ambiguous results in terms of the estimated MeD-set. Although, we discuss only the types of ambiguities occurring in the 2×2, 2×3 and 3×3 cases, we would expect in designs larger than 3×3 additional types of ambiguities occurring and needing resolutions.

In the 2×2 case, there is only one such situation, when $H_0^{(3)}$ is rejected, indicating that there is at least one MeD among the combinations $(1,1)$, $(1,2)$, $(2,1)$ but both null hypotheses $H_0^{(2')}$ and $H_0^{(2'')}$ are accepted, providing contradictory evidence that there is no positive gain combinations among $(1,1)$, $(1,2)$ and among $(1,1)$, $(2,1)$. We call this type of contradiction, *i.e.*, the case when a null hypothesis is tested and rejected and all hypotheses of the lower level implied by it are accepted, Type A ambiguity.

Such ambiguities are due to the constructing a closure of the hypothesis family in the attempt to use the MPG closed testing approach, and therefore, to guarantee the strong control of the FWE. In the 2×2 case, for example, we have to include a redundant hypothesis $H_0^{(3)}$ in the family, because it corresponds to the intersection of $H_0^{(2')}$ and $H_0^{(2'')}$ and makes the family of hypotheses closed. As a consequence, the testing procedure must take into account the decision based on that redundant hypothesis. Other researchers who work on constructing closed testing procedures by including additional intersections of the hypotheses deal with the same problem. For example, Hellmich and Lehmacher (2005) provide an "implication" tree, which specifies the order in which the hypotheses are to be tested for the 2×2 case. This tree is similar to the one presented in Figure 4 but has only 4 hypothesis levels. Nevertheless, they have the same problem of making a decision in terms of the $MED_{<}$-estimate in the case when two individual hypotheses are accepted but the corresponding intersection-hypothesis is rejected.

While the stepwise procedures discussed by Tamhane *et al.* (1996) never result in ambiguities, the problem of dealing with inconclusive results arises even in the single drug case when additional hypotheses are included in the original hypothesis family. Rom *et al.* (1994) consider a problem of identifying the effective doses, *i.e.*, the doses with the response higher than the placebo response. Rom *et al.* (1994) construct a closure of the hypothesis family and implement the additional testing rules (other than the ones specified by the MPG principle), which allow retaining some of the hypotheses without testing. Even under their testing scheme, there are "decision" sequences that lead to contradictory results in terms of identifying the effective doses.

The hypothesis family considered by Tamhane *et al.* (1996) corresponds to the simplest case of an "implication" tree and the minimum effective dose is estimated based on the last rejected hypothesis. In order to provide an estimate in the case of Rom *et al.* (1994) the "decision" sequence needs to be considered and the corresponding conclusions about dose effectiveness need to be made.

In order to eliminate the Type A ambiguity, we can make the conclusion about the estimated MeD-set based on the hypotheses which produce no contradiction. For example, in the 2×2 case since $H_0^{(4)}$ is rejected we believe that there is an MeD and we estimate it as the highest possible combination, *i.e.*, the combination $(2,2)$. Unfortunately, this approach can overestimate the true MeD, so we plan to deal with this case separately from the other cases when the reasonable MeD-set estimate is produced by the procedure. We plan to analyze that type of inconsistency using the simulation study presented in Chapter 6.0.

In the 2×3 case the Type A result also occurs (see Table 8); however, there are "decision" sequences that lead to another type of ambiguity. This kind of ambiguity, Type B, happens, when two hypotheses of the same level are tested, one is accepted and the other is rejected and the hypothesis implied by the rejected one (of the lower level) is tested and accepted. For example, in the case where we deal with rejecting of $H_0^{(6)} : \theta_{11} = \theta_{12} = \theta_{13} = \theta_{21} = \theta_{22} = \theta_{23} = 0$, we test $H_0^{(5)} : \theta_{11} = \theta_{12} = \theta_{13} = \theta_{21} = \theta_{22} = 0$. If it is significant, then we test both $H_0^{(4')} : \theta_{11} = \theta_{12} = \theta_{13} = \theta_{21} = 0$ and $H_0^{(4")} : \theta_{11} = \theta_{12} = \theta_{21} = \theta_{22} = 0$. Suppose now that $H_0^{(4')}$ is rejected and $H_0^{(4")}$ is accepted. Then the MPG principle states that the next hypothesis to be tested is $H_0^{(3')} : \theta_{11} = \theta_{12} = \theta_{13} = 0$ ($H_0^{(3")}$ is not tested because one of the

hypotheses, which implies $H_0^{(4'')}$, is not rejected). The problem arises when we test and accept $H_0^{(3')}: \theta_{11} = \theta_{12} = \theta_{13} = 0$, because it contradicts the conclusion $0 = \theta_{11} = \theta_{12} = \theta_{21} = \theta_{22} < \theta_{13}$, based on the rejected $H_0^{(4')}$ and accepted $H_0^{(4'')}$. There are two "decision" sequences in terms of accepting and rejecting hypotheses that lead to Type B ambiguity and they are noted in Table 8.

To handle the Type B ambiguity we propose to modify the MPG principle, so the testing procedure never results in a Type B situation. If at least one hypothesis of a certain level is accepted, we continue testing only hypotheses of that level and stop testing. The MeD-set is estimated accordingly. In the case considered above, since $H_0^{(5)}$ and $H_0^{(4')}$ are rejected and $H_0^{(4'')}$ is accepted, we could conclude that $0 = \theta_{11} = \theta_{12} = \theta_{21} = \theta_{22} < \theta_{13}$, *i.e.*, the MeD-set is given by $\{(1,3)\}$ without further testing of $H_0^{(3')}$. These ideas are summarized in the "Modified MPG Principle" given below.

Modified MPG principle. Test and reject any hypothesis at level α if and only if all hypotheses implying it, including itself, are significant at level α and all hypotheses of the higher hypotheses levels are significant at level α.

It is straight forward to show that the modified MPG principle leads to strong control of the FWE. Note that the probability of rejecting of at least one true hypothesis under the modified MPG principle is at most the probability of rejecting of at least one true hypothesis under the original MPG principle (because the modified procedure rejects less often). Since the MPG guarantees the strong control of the FWE, the modified principle also guarantees the strong control of the FWE.

Even more complicated cases arise in the 3×3 case. Testing under the modified MPG principle stated for the 2×3 case eliminates Type B ambiguity. However, in the 3×3 case, the MPG principle introduces a new ambiguity, Type C, and while the modified MPG method again eliminates Type B ambiguities, it does not change Type C ambiguity. All possible "decision" sequences for the 3×3 case are summarized in the Table 9. If there are three hypotheses of the same level are tested and exactly two of them are accepted, then Type C ambiguity happens if one of the following is true:

1) The two accepted hypotheses are not implied by the same rejected hypothesis of the higher level,

Table 9. "Decision" sequences (under the modified MPG principle) leading to ambiguities for 3 by 3 case

(For some "decision" sequences there are partial ambiguities while some of the MeD's are identified)

$H_0^{(9)}$	REJ	REJ	REJ	REJ	REJ	REJ	REJ	REJ	REJ
$H_0^{(8)}$	REJ	REJ	REJ	REJ	REJ	REJ	REJ	REJ	REJ
$H_0^{(7')}$	ACC	REJ	REJ	REJ	REJ	REJ	REJ	REJ	REJ
$H_0^{(7'')}$	ACC	REJ	REJ	REJ	REJ	REJ	REJ	REJ	REJ
$H_0^{(6')}$	NT	ACC	ACC	ACC	REJ	REJ	REJ	REJ	REJ
$H_0^{(6'')}$	NT	ACC	ACC	REJ	ACC	REJ	REJ	REJ	REJ
$H_0^{(6''')}$	NT	ACC	REJ	ACC	ACC	REJ	REJ	REJ	REJ
$H_0^{(5')}$	NT	NT	NT	NT	NT	ACC	ACC	ACC	REJ
$H_0^{(5'')}$	NT	NT	NT	NT	NT	ACC	ACC	REJ	ACC
$H_0^{(5''')}$	NT	NT	NT	NT	NT	ACC	REJ	ACC	ACC
$H_0^{(4')}$	NT	NT	NT	NT	NT	NT	NT	NT	NT
$H_0^{(4'')}$	NT	NT	NT	NT	NT	NT	NT	NT	NT
$H_0^{(4''')}$	NT	NT	NT	NT	NT	NT	NT	NT	NT
$H_0^{(3')}$	NT	NT	NT	NT	NT	NT	NT	NT	NT
$H_0^{(3'')}$	NT	NT	NT	NT	NT	NT	NT	NT	NT
$H_0^{(3''')}$	NT	NT	NT	NT	NT	NT	NT	NT	NT
$H_0^{(2')}$	NT	NT	NT	NT	NT	NT	NT	NT	NT
$H_0^{(2'')}$	NT	NT	NT	NT	NT	NT	NT	NT	NT
$H_0^{(1)}$	NT	NT	NT	NT	NT	NT	NT	NT	NT
$\hat{M}$	Type A	Type A	Type A, (2,3)	Type C	Type A, (3,2)	Type A	Type C	Type C	Type C

Table 9 ***continued***

$H_0^{(9)}$	REJ	REJ	REJ	REJ	REJ	REJ	REJ	REJ	REJ
$H_0^{(8)}$	REJ	REJ	REJ	REJ	REJ	REJ	REJ	REJ	REJ
$H_0^{(7')}$	REJ	REJ	REJ	REJ	REJ	REJ	REJ	REJ	REJ
$H_0^{(7'')}$	REJ	REJ	REJ	REJ	REJ	REJ	REJ	REJ	REJ
$H_0^{(6')}$	REJ	REJ	REJ	REJ	REJ	REJ	REJ	REJ	REJ
$H_0^{(6'')}$	REJ	REJ	REJ	REJ	REJ	REJ	REJ	REJ	REJ
$H_0^{(6''')}$	REJ	REJ	REJ	REJ	REJ	REJ	REJ	REJ	REJ
$H_0^{(5')}$	REJ	REJ	REJ	REJ	REJ	REJ	REJ	REJ	REJ
$H_0^{(5'')}$	REJ	REJ	REJ	REJ	REJ	REJ	REJ	REJ	REJ
$H_0^{(5''')}$	REJ	REJ	REJ	REJ	REJ	REJ	REJ	REJ	REJ
$H_0^{(4')}$	ACC	ACC	ACC	REJ	REJ	REJ	REJ	REJ	REJ
$H_0^{(4'')}$	ACC	REJ	ACC	ACC	REJ	REJ	REJ	REJ	REJ
$H_0^{(4''')}$	ACC	ACC	REJ	ACC	REJ	REJ	REJ	REJ	REJ
$H_0^{(3')}$	NT	NT	NT	NT	ACC	ACC	ACC	REJ	REJ
$H_0^{(3'')}$	NT	NT	NT	NT	ACC	ACC	REJ	ACC	REJ
$H_0^{(3''')}$	NT	NT	NT	NT	ACC	REJ	ACC	ACC	REJ
$H_0^{(2')}$	NT	NT	NT	NT	NT	NT	NT	NT	ACC
$H_0^{(2'')}$	NT	NT	NT	NT	NT	NT	NT	NT	ACC
$H_0^{(1)}$	NT	NT	NT	NT	NT	NT	NT	NT	NT
$\hat{M}$	Type A	Type A, (3,1)	Type A, (1,3)	Type A, (2,2)	Type A	Type A, (1,3) (2,2)	Type A, (3,1) (2,2)	Type A, (1,2)	Type A

2) All three hypotheses are implied by the same rejected hypothesis of the higher level.

For example, if we reject all hypotheses of the 6th level, accept $H_0^{(5')}: \theta_{11} = \theta_{12} = \theta_{21} = \theta_{22} = \theta_{31} = 0$, reject $H_0^{(5'')}: \theta_{11} = \theta_{12} = \theta_{13} = \theta_{21} = \theta_{31} = 0$ and accept $H_0^{(5''')}: \theta_{11} = \theta_{12} = \theta_{13} = \theta_{21} = \theta_{22} = 0$ then we stop. In this case all three hypotheses of the 5th level follow from the same hypothesis $H_0^{(6'')}$ so this is the Type C ambiguity. Acceptance of $H_0^{(5')}$ and $H_0^{(5''')}$ leads to the conclusion that none of the combinations $(1,1)$, $(1,2)$, $(1,3)$, $(2,1)$, $(2,2)$, $(3,1)$ belongs to the MeD-set but rejection of $H_0^{(5'')}$ indicates that at least one of $(1,1)$, $(1,2)$, $(1,3)$, $(2,1)$, $(3,1)$ belongs to the MeD-set. Thus, no meaningful conclusion in terms of the estimated MeD-set can be made.

The Type C ambiguity is certainly different from the Type A inconclusive result. In the example above, Type A ambiguity would occur if all hypotheses of the 5th level are accepted, because Type A ambiguity occurs if a certain hypothesis is rejected but all hypotheses of the lower level implied by it are accepted.

Also in the 3×3 case some "decision" sequences lead to the combination of the reasonable MeD-estimate together with Type A inconclusive result (see Table 9). For example, suppose we reject both hypotheses of the 7th level, $H_0^{(7')}: \theta_{11} = \theta_{12} = \theta_{13} = \theta_{21} = \theta_{22} = \theta_{31} = \theta_{32} = 0$ and $H_0^{(7'')}: \theta_{11} = \theta_{12} = \theta_{13} = \theta_{21} = \theta_{22} = \theta_{23} = \theta_{31} = 0$, and we accept both $H_0^{(6')}: \theta_{11} = \theta_{12} = \theta_{21} = \theta_{22} = \theta_{31} = \theta_{32} = 0$ and $H_0^{(6'')}: \theta_{11} = \theta_{12} = \theta_{13} = \theta_{21} = \theta_{22} = \theta_{32} = 0$, and reject $H_0^{(6''')}: \theta_{11} = \theta_{12} = \theta_{13} = \theta_{21} = \theta_{22} = \theta_{23} = 0$. Since both accepted hypotheses $H_0^{(6')}$ and $H_0^{(6'')}$ are implied by rejected $H_0^{(7')}$ and they are the only ones implied by it, Type A ambiguity occurs. At the same time, based on acceptance of $H_0^{(6'')}$ and rejection of $H_0^{(6''')}$ and $H_0^{(7'')}$, we can conclude that $(2,3)$ belongs to the MeD-set. Although, in this case we can not provide a proper estimate for the whole MeD-set, there is enough evidence to believe that a combination dose $(2,3)$ belongs to the MeD-set.

To summarize the above results, we can claim that in the considered designs a Type A or Type C ambiguity happens if and only if at least two hypotheses of the same level are accepted. We use the simulation studies to approximate the probabilities of occurring Type A ambiguity

for 2×2 and 2×3 designs. We also consider testing under the modified MPG principle for the 2×3 design and provide the results in Chapter 6.0, where the simulation studies are discussed.

Again, we would expect in designs larger than 3×3 additional types of ambiguities occurring and needing resolution.

5.0 CRITERIA FOR EVALUATING TESTING PROCEDURES

5.1 ON PARAMETER ESTIMATION

In the case of a single drug study with K active doses, we consider the problem of identifying an MED. The following facts are true about the MED in a single drug case:

1) The population parameter of interest, the MED, has a simple structure; if the MED exists then it is unique and is a single element.
2) Under the assumption of non-decreasing mean dose-response function, the MED provides the full information about the drug effectiveness. All doses to the left of the MED are non-effective and all doses to the right are effective.
3) If the dose-response is not assumed to be monotone, then the MED identifies the effectiveness of only lower doses than the MED. All doses with lower dose levels than the MED are identified as non-effective.
4) The performance of the testing procedure can be measured by using Familywise Error, Power, Lack of Power and Bias. These values can be easily obtained by simulation.

Now let us consider the analogs of the above statements for the case of the $K \times N$ combination drug study regarding the MeD-set.

1) The population parameter of interest is the MeD-set. If it exists (non-empty) then it is unique, but it may have a very complicated structure.

   The cardinality of the set can be any number from 0 to $\min\{K, N\}$. Moreover, the single element of the set does not provide the information about other possible elements of the MeD-set, unless it is the smallest combination, (1, 1), or the largest combination, (K, N). For example, for the 3×3 design, if (1, 3) belongs to the MeD-set then it does not specify the cardinality of the set (it can be 1, 2 or 3) and it does not provide any information about other combinations that, possibly, belong to the set. We can only

conclude that the MeD-set is one of the following four sets:

{(1, 3)}, {(1, 3), (2, 2)}, {(1, 3), (3, 1)}, {(1, 3), (2, 2), (3, 1)}.

2) Under the assumption of an isotonic structure of the indicator gain matrix the MED provides the full information about the drug effectiveness in terms of the combination gains. All doses with at least one index smaller than the MeD's are no-gain combinations and all doses with both indexes bigger than the MeD's are positive gain combinations.
3) If a general structure of an indicator gain matrix is assumed, then MeD's provide the information only about combinations with at least one index smaller than the MeD's; all such combinations are identified as no-gain doses.
4) A performance of the testing procedure can not be measured using the same notions as in a single drug case. The measures such as the Familywise Error, Power, Lack of Power and Bias must be generalized to the combination drug setting.

Section 5.2 illustrates how the Familywise Error, Power and Lack of Power can be generalized to measure the performance of the testing procedures in the combination drug setting, and how to obtain these by simulations.

While Bias is well defined for a single drug study, the problem of defining it for the combination drug design is not trivial. It lies in the definition of the expected value of the estimator and measuring the distance from this expectation to the true parameter, the MeD-set. Hellmich and Lehmacher (2005) just mention that the notion of Bias can be extended to the combination drug case but do not provide any Bias-like measure.

Intrinsically, to measure bias, one has to compare two sets of, perhaps, different cardinality. Instead of the Bias, we consider the loss functions defined in Section 5.3 which can be used to measure the "distance" between the estimated MeD-set and the true MeD- set.

In addition to the measures described above we define two notions which are related to overestimating and underestimating the population MeD-set. These notions are discussed in Section 5.4.

5.2 FAMILYWISE ERROR, POWER AND LACK OF POWER

In this section we illustrate how to measure the performance of testing procedures constructed to identify the MeD-set in terms of the estimated Familywise Error (FWE), Power and Lack of Power (LOP). These measures can be used in the combination drug case even when the isotonic structure of the indicator gain matrix is not assumed.

First, we provide the statistical definitions of the above measures and then discuss the rules for estimating these measures by simulation. We use the following definitions which are suitable for assessing performance of any closed testing procedure. Familywise Error is defined as in a single drug case, *i.e.*, it is the probability of rejecting at least one true null hypothesis. Power is defined as the probability of rejecting the set of all false and only false null hypotheses. Lack of Power is the probability of accepting at least one false null hypothesis. Care must be taken in specifying the parameter configurations under which these probabilities are computed.

Following Tamhane *et al.* (1996) we measure the performance of an MED estimation procedure in the case of a single drug using the above notions. In the single drug case these measures were obtained by simulation and applying formulas (2.25). The proportion of replications for which a lower dose than the true MED is estimated as an MED corresponds to the estimated Familywise Error of the procedure. The proportion of replicates for which the MED is estimated correctly provides the estimated Power of the procedure. The Lack of Power is estimated by the proportion of cases when the doses with higher indexes than the true MED's are estimated as the MED's including the cases when none of the doses is declared as the MED. Note, that in the single drug case, we can write that FWE is the probability that lower doses than the true MED is estimated as an MED. Power is the probability that the MED is estimated correctly. And LOP is the probability that the doses with higher indexes than the true MED's are estimated as the MED's.

First, let us illustrate why the rules for obtaining the simulated FWE, Power and LOP need to be generalized for the combination drug setting. Let us consider a simple example, when the true MeD-set is given by $\{(1,2)\}$ and it is estimated as $\{(2,1)\}$. Since these two doses are incomparable, it is not clear whether the estimate provides a higher or a lower dose combination than the true MeD, so some modifications are needed in order to use these measures.

Hellmich and Lehmacher (2005) use some modifications of these notions to measure the performance of their procedure to identify the minimum effective doses (MED's) in the combination drug case. The authors consider the combination drug design under the assumption of an isotonic structure of the mean response matrix, which makes it possible for them to classify each combination as non-effective, effective but not minimum effective, or as an MED. Using this information, they simulate the FWE, Power and LOP using the following statements: FWE is given by the proportion of cases when non-effective doses are estimated as the MED's. Power is obtained as the proportion of replications corresponding to the correctly estimated MED's. Lack of Power is the proportion of cases when effective doses but not MED's are estimated as the MED's.

We propose a generalization of the estimated measures used by Hellmich and Lehmacher (2005). We obtain the estimated measures by utilizing all available information about dose effectiveness provided by the true and estimated MeD-sets. The advantage of such a reformulation is that it introduces these concepts to situations where the isotonic structure of the gain matrix is not assumed.

In order to simulate the values of our generalizations of the estimated FWE, Power and Lack of Power, we consider the following proportions. The proportion of replications for which at least one no-gain dose is estimated as an MeD corresponds to the estimated FWE of the procedure. In this case at least one true MeD is underestimated. The proportion of replicates for which the exact MeD-set is estimated correctly provides the estimate of the Power of the procedure, including the cases when an empty MeD-set is estimated correctly as an empty one. The Lack of Power is estimated by the proportion of all cases when a combination with unknown true effectiveness is estimated as the MeD, including the cases when none of the combinations is declared as the MeD but the true MeD-set is not empty. If the population MeD-set is non-empty and it is estimated as an empty set, then all combinations are estimated as no-gain doses, meaning that the MeD, if it exists, has higher indexes than the doses considered in the study, hence, at least one true MeD is overestimated.

Next, we provide an approach to translate the FWE, Power and LOP in terms of probabilities involving the true and estimated MeD-sets.

Let the population MeD-set, M , and the MeD-set, $\hat{M}$, estimated by a statistical procedure be denoted by

$$M \equiv \begin{cases} \{c_s = (i_s, j_s),\ s = 1,...,S\}, & for\ 1 \le S \le \min\{K, N\} \\ \varnothing, & if\ M\ is\ empty \end{cases} \tag{5.1a}$$

and

$$\hat{M} \equiv \begin{cases} \{c'_t = (i'_t, j'_t),\ t = 1,...,T\}, & for\ 1 \le T \le \min\{K, N\} \\ \varnothing, & if\ \hat{M}\ is\ empty, \end{cases} \tag{5.1b}$$

respectively, where $1 \le i, i' \le K$, $1 \le j, j' \le N$. If the MeD-set is empty then there are no efficacious combinations in the grid $K \times N$, but there may be efficacious doses in the larger grid so it makes more sense to use $\{(K+1, N+1)\}$ to represent the empty MeD-set than $\{(0,0)\}$.

Next, let M^- denote a set of all combinations which are identified by the MeD-set as no-gain combinations. These are the combinations with at least one index lower than the true MeD's. And let M^+ denote a set of all combinations with both indexes higher than the true MeD's. This latter set may consist of efficacious and/or no-gain combinations if the isotonic structure of the population indicator gain matrix is not assumed. If the isotonicity is assumed then the set M^+ consists of efficacious doses. Based on the given MeD-set each element (i, j) , $i = 1,2,...,K,\ \ j = 1,2,...,N$ of the matrix Θ' can be allocated to one and only one of the sets: MeD-set, M , no-gain dose set, M^-, or the set of all other combinations, M^+.

Similarly, based on the estimated MeD-set, $\hat{M}$, and the corresponding indicator gain matrix, $\hat{\Theta}'$, each combination (i, j) , $i = 1,2,...,K,\ \ j = 1,2,...,N$ can be allocated to one and only one of the sets: estimated MeD-set, $\hat{M}$, estimated no-gain dose set, $\hat{M}^-$, or the estimated set $\hat{M}^+$, which contains the combinations with higher dose indexes than the estimated MeD's.

Using the preceding notation, we rewrite the definitions of the FWE, Power and Lack of Power as follows.

Statistical Measure	Statistical Definition	Alternative Definition
Familywise Error (FWE)	Probability of rejecting at least one true null hypothesis	$P(\hat{M} \cap M^{-} \neq \varnothing)$
Power	Probability of rejecting all false and only false null hypotheses	$P(\hat{M} = M)$
Lack of Power (LOP)	Probability of accepting at least one false null hypothesis	$P(\hat{M} \cap M^{+} \neq \varnothing)$

Note that measures defined in such a way add up to 1. In the case when a closed testing procedure results in ambiguity, the alternative definitions can not be used, but if each type of an ambiguity corresponds to a single and known "decision" sequence then the statistical definitions can be used in order to obtain the FWE, Power and LOP. Such situations are discussed in detail later on in this section. In the cases when a procedure results in some MeD-set estimate, it is more convenient to use the alternative statements.

We do not provide an explicit proof that these definitions are equivalent other than in the 2×2 and 2×3 cases, as they involve tedious enumerations. Our judgment is that these definitions would extend to the $K \times N$ case, but we have not yet obtained the necessary complex graph-theoretic proof.

Now let us come back to the above example when the true MeD-set is given by $\{(1,2)\}$ and it is being estimated as $\{(2,1)\}$. Then this case results in FWE, because a no-gain combination $(2,1)$ is estimated as the MeD.

Some procedures may produce no meaningful MeD-set estimate. Then these cases should be considered separately in detail to figure out what kind of error this outcome corresponds to. For example, the proposed step-down procedure can result in some type of ambiguity. Let us consider the case when it results in Type A ambiguity. Other types can be considered similarly.

The proportion of cases when Type A ambiguity happens, should be included in calculating the FWE or LOP. In the case of Type A ambiguity our procedure stops due to acceptance of at least two hypotheses of some level, both implied by the same rejected hypothesis. For example, this happens in the 2×2 case, when $H_0^{(3)}$ is rejected but both hypotheses $H_0^{(2')}$ and $H_0^{(2'')}$ are accepted. These situations should be analyzed in more detail,

because depending on the true responses they may lead to the different types of error committed by the testing procedure. For example, if the true gains are $\theta_{11} = \theta_{12} = \theta_{21} = 0$ then the proportion of cases when Type A ambiguity happens contributes to the FWE, because the procedure fails to reject the true $H_0^{(3)}$. If the true gains are not all equal to zero, then the procedure rejects the false $H_0^{(3)}$ (no error made) but it fails to reject at least one false hypothesis ($H_0^{(2')}$ or $H_0^{(2'')}$). The last corresponds to the LOP and the proportion of such cases should be allocated to this measure.

Example 5.1. Suppose, for the 2×2 case the true MeD-set is given by $\{(1,2)\}$. Then the population MeD-matrix and the population indicator gain matrix are given by

$$\Lambda = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \text{ and } \Theta' = \begin{pmatrix} 0 & 1 \\ 0 & * \end{pmatrix}.$$

Suppose that the simulation studies based on 10,000 replications produce the results given in Table 10. Then $Power = 8{,}000/10{,}000 = 0.8$. Combinations $(1,1)$ and $(2,1)$ are known to be no-gain combinations, so $FWE = (200 + 200 + 300)/10{,}000 = 0.07$.

Table 10. Estimated MeD-set with the corresponding hypothetical frequency

Result in Terms of Estimated MeD-set	Frequency
Empty	600
$\{(2,2)\}$	400
Type A: $H_0^{(4)}$, $H_0^{(3)}$ are rejected and $H_0^{(2')}$, $H_0^{(2'')}$ are accepted	300
$\{(2,1)\}$	200
$\{(1,2)\}$	8,000
$\{(2,1), (1,2)\}$	200
$\{(1,1)\}$	300

Next, the Type A result indicates that the false null hypotheses $H_0^{(4)}$ and $H_0^{(3)}$ are rejected (no error made), the true $H_0^{(2'')}$ is accepted (no error made), but the false $H_0^{(2')}$ is also

accepted (type II error made). So the replications corresponding to this case should be included in the computations of Lack of Power. Also, the proportion of cases when the estimated MeD-set is given by a single combination $(2,2)$ together with the replications corresponding to the "Empty" outcome should contribute to the Lack of Power:

$Lack\ of\ \ Power = (600 + 400 + 300)/10{,}000 = 0.13$.

To reiterate, this example is based on hypothetical simulation data and the frequencies are given for pure illustration of the methods for obtaining the simulated FWE, Power and Lack of Power.

5.3 LOSS FUNCTION

To measure the performance of an estimator of the MeD-set, we define several loss functions and describe their properties. Let the population MeD-set and the MeD-set estimated by a statistical procedure be denoted by (5.1) of Section 5.2. In order to define a loss function $L(M,\ \hat{M})$ we use the key feature of any loss function: $L(M,\ \hat{M}) \in [0,\infty)$ with the property $L(M,\ M) = 0$.

We propose several definitions of the loss functions based on the distances between the dose combinations that belong to M and the dose combinations that belong to $\hat{M}$. Since both sets can be represented uniquely by the $K \times N$ MeD-matrices, we can introduce the notion of distance based on the relative locations of two combinations in the $K \times N$ matrix. By the distance, $d(a_1, a_2)$, between two dose combinations $a_1 = (p_1, q_1)$ and $a_2 = (p_2, q_2)$, where $p_1, p_2 \in \{1,2,...,K\}$, $q_1, q_2 \in \{1,2,...,N\}$, we mean $d(a_1, a_2) = |p_1 - p_2| + |q_1 - q_2|$. Note that if $a_1 = a_2$ (*i.e.*, $p_1 = p_2$ and $q_1 = q_2$), then $d(a_1, a_2) = 0$. We also use the notion of the distance between a dose combination $a^* = (p^*, q^*)$ with $p^* \in \{1,2,...,K\}$, $q^* \in \{1,2,...,N\}$ and a set $A = \{a_i = (p_i, q_i),\ 1 \le i \le \min\{K, N\}\}$, defined as $d^*(a^*, A) = \min_{all\ i} d(a^*, a_i)$. Note that if $a^* \in A$, then $d^*(a^*, A) = 0$.

Definition 5.1 for the loss function does not require additional knowledge of the drug combinations other than whether or not they are included in the given MeD-sets and is given below.

Definition 5.1. Let M and $\hat{M}$ denote, respectively, the true and estimated MeD-sets. Further, let $M_1 \equiv \hat{M} - \left(M \cap \hat{M}\right)$ and $M_2 \equiv M - \left(M \cap \hat{M}\right)$. Then the loss function, L_1, is defined as

$$L_1(M,\ \hat{M}) = \sum_{c' \in M_1} d^*(c',\ M) + \sum_{c \in M_2} d^*(c,\ \hat{M}).$$

In the case when $\hat{M}$ is not identical to M, each element $c' \notin M$ increases the value of the first sum, and each element $c \notin \hat{M}$ increases the value of the second sum in the expression for $L_1(M,\ \hat{M})$.

Example 5.2. For simplicity of calculations, we consider the 2×2 design with the true MeD-set $M = \{(1,2)\}$ and two estimated sets, $\hat{M}_1 = \{(2,2)\}$ and $\hat{M}_2 = \{(1,1)\}$. Let us compare these estimates based on the values of the loss functions given in Definition 5.1. Since $M \cap \hat{M}_i$, $i = 1,2$ is empty then $M_1 \equiv \hat{M}_i$, $i = 1,2$ and $M_2 \equiv M$ in both cases. So

$$L_1(M,\ \hat{M}_1) = L_1(M,\ \hat{M}_2) = 1 + 1 = 2.$$

Next, let us look at the drawbacks of the loss function, defined above. First of all, $L_1(M,\ \hat{M})$ does not take into account whether the combination, estimated as an MeD is indeed an efficacious (but not minimum efficacious) combination or it is a no-gain combination. Example 5.2 illustrates that the estimators $\hat{M}_1 = \{(1,1)\}$ and $\hat{M}_2 = \{(2,2)\}$ have the equal loss functions, although, the second estimator provides either an efficacious or no-gain dose as an MeD-estimate, while the first estimator provides a no-gain dose $(1,1)$ as an MeD. So if no more information is given about the other combinations, it is logical to have $L(M,\ \hat{M}_1) \neq L(M, \hat{M}_2)$.

Whether we want $L(M,\ \hat{M}_1) > L(M, \hat{M}_2)$ or $L(M,\ \hat{M}_1) < L(M, \hat{M}_2)$ depends on the particular drugs under study. If one prefers overestimating an MeD to underestimating it, then it is desirable to have $L(M,\ \hat{M}_1) > L(M, \hat{M}_2)$. This situation can arise when the administration of the higher drug dose combinations is not harmful and it is more important to see the effect than to control the amount of the administered drug. If administering the higher drug doses is not safe then one favors underestimating an MeD, rather than overestimating it. Hence, it is preferable to

have $L(M,\ \hat{M}_1) < L(M, \hat{M}_2)$. The loss function given in Definition 5.1 does not take into account those different cases and treats them equally.

More flexible loss functions can be constructed in the case when additional information about the combination effectiveness is taken into account.

Table 11. Possible allocations of a single dose with the corresponding errors

Population Property of a Dose (i, j)	Estimated Property of a Dose (i, j)		
	$\hat{M}^-$	$\hat{M}$	$\hat{M}^+$
M^-	No Error	Error C*	Error D*
M	Error B*	No Error	Error A*
M^+	Error F*	Error E*	No Error

Using the notation introduced in Section 5.2, we can allocate each dose (i, j), $i = 1,2,...,K$ and $j = 1,2,...,N$ to one and only one set based on its true and estimated property. Table 11 summarizes the possible dose allocations with the corresponding conclusions whether or not a statistical procedure makes an estimation error. The table uses the following notation.

By "No Error" for dose (i, j), we mean the correct estimation. "No Error" corresponds to the cases when there is no information available about how a combination with unknown true effectiveness is estimated. And since it is correctly allocated to $\hat{M}^+$ set, the procedure does not make an estimation error in terms of the dose allocation.

Now, let us consider the cases when the estimation of the dose effectiveness leads to an error. There are six types of errors. We say that A* error occurs when $(i, j) \in M \cap \hat{M}^+$, in this case the true MeD is underestimated and as the result, it is estimated as an $\hat{M}^+$-dose. We say that B* error occurs when $(i, j) \in M \cap \hat{M}^-$, *i.e.*, the effectiveness of the true MeD is not detected by a procedure and it is estimated as a no-gain combination. In this case the MeD is either overestimated or the estimated MeD-set is empty. C* error corresponds to the case when

$(i, j) \in M^- \bigcap \hat{M}$, *i.e.*, a no-gain combination is estimated as an MeD, so all lower no-gain combinations are estimated correctly. D* error occurs when $(i, j) \in M^- \bigcap \hat{M}^+$. In this case the true MeD is misclassified. If we compare the cases related to last two errors then there is the larger number of misclassifications associated with the error D*. Finally, if $(i, j) \in M^+ \bigcap \hat{M}$ or $(i, j) \in M^+ \bigcap \hat{M}^-$ then the true MeD is overestimated and these cases denoted by errors F* and E*, respectively. There are more misclassifications happening in a case resulting in F* error than resulting in E* error.

The above classification provides a better understanding of the desirable properties of a loss function. Clearly, the different types of errors should have different impact on the value of the loss function and can be weighted depending on the researcher's preference. The following definition of the loss function allows assigning the different values,

w_{A^*}, w_{B^*}, w_{C^*}, w_{D^*} w_{E^*}, w_{F^*}, to these types of errors.

Definition 5.2. For each drug combination (i, j), $i = 1,..., K$, $j = 1,..., N$ consider its unique allocation to one of the classes presented in Table 12 with the assigned value of error w_{ij}, where $w_{ij} \in \{0, w_{A^*}, w_{B^*}, w_{C^*}, w_{D^*}, w_{E^*}, w_{F^*}\}$, $i = 1,..., K$, $j = 1,..., N$ are some fixed positive constants. Then the loss function is defined by $L_2(M, \hat{M}) = \sum_{i=1}^{K} \sum_{j=1}^{N} w_{ij}$.

Table 12. Possible allocations of a single dose with the assigned values of errors

Population Property of a Dose (i, j)	Estimated Property of a Dose (i, j)		
	$\hat{M}^-$	$\hat{M}$	$\hat{M}^+$
M^-	0	w_{C^*}	w_{D^*}
M	w_{B^*}	0	w_{A^*}
M^+	w_{F^*}	w_{E^*}	0

Since the function defined in such a way is equal to zero if and only if all of the errors w_{ij} are zero (*i.e.*, all combinations are estimated correctly), it is a proper definition of the loss function. Obviously, the more the number of incorrectly estimated combinations, the higher the value of $L_2(M,\ \hat{M})$.

Example 5.3 illustrates one of the possible assignments of the values to the constants $w_{A^*}, w_{B^*}, w_{C^*}, w_{D^*}, w_{E^*}, w_{F^*}$.

Example 5.3. Suppose that a dose-response study deals with the drugs such that the higher drug dose combinations are not as safe as the lower ones, and it is preferable to underestimate the MeD than to overestimate it. Among the cases presented in Table 12, there are three cases that deal with overestimating the true MeD (and result in B*, E* or F*error) and three cases that correspond to underestimating the MeD (and result in A*, C* or D* error). Since error F* is the worst in terms of the number of misclassified combinations, and the errors B* and E* are relatively the same, we require $w_{B^*} = w_{E^*} < w_{F^*}$. Similarly, among the cases when the true MeD is underestimated we can order the types of errors as $w_{A^*} = w_{C^*} < w_{D^*}$ and so that we choose the error values such that $w_{A^*} = w_{C^*} < w_{D^*} \le w_{B^*} = w_{E^*} < w_{F^*}$. Using this fact we could use, for example, $w_{A^*} = w_{C^*} = 1,\ w_{D^*} = 1.5,\ w_{B^*} = w_{E^*} = 2,\ w_{F^*} = 2.5$ as one of the possible assignments.

Next, we illustrate how the loss function $L_2(M,\ \hat{M})$ can be calculated based on these error values for the MeD-sets given in Example 5.2. The MeD-sets are given by $M = \{(1,2)\}$, $\hat{M}_1 = \{(2,2)\}$ and $\hat{M}_2 = \{(1,1)\}$. Obviously, $\hat{M}_1$ overestimates the MeD and $\hat{M}_2$ underestimates the true MeD. While $\hat{\Theta}'_1$ is uniquely identified by $\hat{M}_1$, there are a number of different Θ' and $\hat{\Theta}'_2$ corresponding to M and $\hat{M}_2$, respectively. The following are the steps that allow to calculate L_2.

Let m^+ represent a combination whose effectiveness is not uniquely identified by the true MeD-set ($m^+ \in M^+$), or by the estimated MeD-set ($m^+ \in \hat{M}^+$). Then we obtain

$$\Theta' = \begin{pmatrix} 0 & 1 \\ 0 & m^+ \end{pmatrix},\ \hat{\Theta}'_1 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \text{ and } \hat{\Theta}'_2 = \begin{pmatrix} 1 & m^+ \\ m^+ & m^+ \end{pmatrix}.$$

Let us first calculate the loss function corresponding to the $\hat{M}_1$. Based on the first two matrices, each dose combination is classified as follows: $\{(1,1),(2,1)\} \in M^- \bigcap \hat{M}^-$, $(1,2) \in M \bigcap \hat{M}^-$, $(2,2) \in M^+ \bigcap \hat{M}$. Then, using the above assignment of the misclassification errors, we obtain $w_{11} = w_{21} = 0$, $w_{12} = w_{B^*} = 2$, $w_{22} = w_{E^*} = 2$ and, therefore,

$$L_2(M,\ \hat{M}_1) = \sum_{i=1}^{2}\sum_{j=1}^{2} w_{ij} = 4.$$

Next, let us calculate the loss function $L_2(M,\ \hat{M}_2)$. Since $(1,1) \in M^- \bigcap \hat{M}$, $(1,2) \in M \bigcap \hat{M}^+$, $(2,1) \in M^- \bigcap \hat{M}^+$ and $(2,2) \in M^+ \bigcap \hat{M}^+$ with the corresponding errors $w_{11} = w_{C^*} = 1$, $w_{12} = w_{A^*} = 1$, $w_{21} = w_{D^*} = 1.5$ and $w_{22} = 0$, then $L_2(M,\ \hat{M}_2) = 3.5 < L_2(M, \hat{M}_1)$. So $\hat{M}_2$ provides a better estimation than $\hat{M}_1$ in terms of the loss function $L_2(M,\ \hat{M})$. This concludes the example.

Now, let us consider a special case of L_2, where we do not differentiate among all the error types. First of all, any true (estimated) MeD-set defines uniquely both sets, M^- ($\hat{M}^-$) and M^+ ($\hat{M}^+$), and, therefore, we can consider no-gain doses separately from the other doses, *i.e.*, we can consider only two sets, M^- ($\hat{M}^-$) and $U = M \bigcup M^+$ ($\hat{U} = \hat{M} \bigcup \hat{M}^+$). Then, if $(i, j) \in M^- \bigcap \hat{U}$ and the population MeD-set is not empty then at least one true MeD is underestimated. If $(i, j) \in M^- \bigcap \hat{U}$ and the population MeD-set is empty then at least one true no-gain combination is identified as an MeD by the procedure. In any case, the true MeD is underestimated.

If $(i, j) \in U \bigcap \hat{M}^-$ and $\hat{U} = \varnothing$ then all combinations are identified as no-gain combinations. Hence, no MeD is detected by the procedure. If $(i, j) \in U \bigcap \hat{M}^-$ and $\hat{U} \neq \varnothing$ then a larger dose combination than the true MeD is declared as an MeD. In any case, at least one true MeD is overestimated.

All other allocations of a combination (i, j) can be viewed as the correct ones with the corresponding allocation errors being zero.

Table 13. Possible single dose allocation with the corresponding conclusion about the true MeD and the assigned errors

Population Property of a Dose (i, j)	Estimated Property of a Dose (i, j)	
	$\hat{M}^-$	$\hat{U} = \hat{M} \bigcup \hat{M}^+$
M^-	No Error, $w_{ij} = 0$	Underestimated, $w_{ij} = w_u$
$U = M \bigcup M^+$	Overestimated, $w_{ij} = w_o$	No Error, $w_{ij} = 0$

All of the above cases are presented in Table 13. Using this logic we need to assign only two error values, one is for overestimating, w_o, and the other is for underestimating, w_u, the true MeD. Then, the following loss function can be defined.

Definition 5.3. For each drug combination (i, j), $i = 1,...,K$, $j = 1,...,N$ consider its allocation presented in Table 13 with the assigned value of error w_{ij}, where $w_{ij} \in \{0, w_o, w_u\}$ for all $i = 1,...,K$ and $j = 1,...,N$. Then the loss function is defined by $L_3(M, \hat{M}) = \sum_{i=1}^{K} \sum_{j=1}^{N} w_{ij}$.

The loss function defined in such a way involves fewer numerical computations than the one defined in Definition 5.2, and at the same time it allows assigning different misclassification error values. Let us illustrate how $L_3(M, \hat{M})$ can be calculated for the set-up of the Example 5.2.

Example 5.4. Consider the set-up of the Example 5.2 with

$$\Theta' = \begin{pmatrix} 0 & 1 \\ 0 & m^+ \end{pmatrix}, \ \hat{\Theta}'_1 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \text{ and } \hat{\Theta}'_2 = \begin{pmatrix} 1 & m^+ \\ m^+ & m^+ \end{pmatrix}.$$

Let us first calculate the loss function for $\hat{M}_1 = \{(2,2)\}$. Since $(1,2) \in U \bigcap \hat{M}^-$ is the only misclassified combination, $L_3(M, \hat{M}_1) = w_o$. Similarly, for $\hat{M}_2 = \{(1,1)\}$ we have $\{(1,1),(2,1)\} \in M^- \bigcap \hat{U}$ so $L_3(M, \hat{M}_2) = 2w_u$. If the values of w_o and w_u are specified then the

loss functions can be compared. If we choose $w_o = w_u = 1$ then $L_2(M, \hat{M}_1) = 1 < L_2(M, \hat{M}_2) = 2$. Based on these values, one would rather use a procedure resulting in the first MeD-set. This concludes the example.

We have considered three definitions of the loss functions which can be used to measure the distance between the true MeD-set and the estimated one and, therefore, to allow comparison among estimated MeD-sets. As Examples 5.2-5.4 show, the same estimated MeD-sets can result in different values of the corresponding loss functions, depending on what definition of the loss function is used. Moreover, which estimator is preferred depends on the available information and assumptions that can be made.

While Definition 5.1 does not require any information about the dose combinations other than the true MeD's and the estimated MeD's, Definitions 5.2 and 5.3 use all available information about possible structure of the indicator gain matrix. Definitions 5.2 and 5.3 provide some flexibility in assigning the misclassification errors and can be easily modified to handle different cases. In general, these loss functions are not symmetric with respect to the true and estimated MeD-sets, while the first loss function is symmetric. In the special case when $w_o = w_u$, the third loss function becomes also symmetric.

Table 14 and Table 15 present the values of the first and third loss functions for all configurations of the true and estimated MeD-sets in the 2×2 case. Both loss functions are used in the simulation studies as an analog of Bias to measure the performance of the proposed procedures. These simulation studies are described in Chapter 6.0.

Table 14. Values of the loss function $L_1(M,\ \hat{M})$ for all configurations of the true and estimated MeD-sets in the 2×2 case

M \ $\hat{M}$	Empty	{(2,2)}	{(2,1)}	{(1,2)}	{(2,1), (1,2)}	{(1,1)}
Empty	0	4	6	6	9	8
{(2,2)}	4	0	2	2	3	4
{(2,1)}	6	2	0	4	2	2
{(1,2)}	6	2	4	0	2	2
{(2,1), (1,2)}	9	3	2	2	0	3
{(1,1)}	8	4	2	2	3	0

Table 15. Values of the loss function $L_3(M,\ \hat{M})$ for all configurations of the true and estimated MeD-sets in the 2×2 case

M \ $\hat{M}$	Empty	{(2,2)}	{(2,1)}	{(1,2)}	{(2,1), (1,2)}	{(1,1)}
Empty	0	w_u	$2w_u$	$2w_u$	$3w_u$	$4w_u$
{(2,2)}	w_o	0	w_u	w_u	$2w_u$	$3w_u$
{(2,1)}	$2w_o$	w_o	0	$w_o + w_u$	w_u	$2w_u$
{(1,2)}	$2w_o$	w_o	$w_o + w_u$	0	w_u	$2w_u$
{(2,1), (1,2)}	$3w_o$	$2w_o$	w_o	w_o	0	w_u
{(1,1)}	$4w_o$	$3w_o$	$2w_o$	$2w_o$	w_o	0

5.4 ADDITIONAL MEASURES TO ASSESS PERFORMANCE OF TESTING PROCEDURES

Next, we define two additional measures that characterize the estimate in terms of overestimating and underestimating the true MeD-set. Let the true MeD-set and the estimated one be represented by (5.1).

We say that the population MeD-set $M = \{c_s\}$, is underestimated by $\hat{M} = \{c_t'\}$ if for any element $c_t' \in \hat{M}$ there exists an element $c_s \in M$ such that $c_t' \leq c_s$ with at least one strict inequality.

Similarly, we say that the population MeD-set $M = \{c_s\}$ is overestimated by $\hat{M} = \{c_t'\}$ if for any element $c_s \in M$ there exists an element $c_t' \in \hat{M}$ such that $c_s \leq c_t'$ with at least one strict inequality.

For example, in the 2×3 case if the true MeD-set is given by $M = \{(2,2),\ (1,3)\}$ then it is underestimated by each of the following sets: $\{(2,1)\}$, $\{(1,2)\}$, $\{(2,1),\ (1,3)\}$, $\{(2,1),\ (1,2)\}$, $\{(1,1)\}$. And it is overestimated by an empty set, $\{(3,4)\}$, and by a set $\{(2,3)\}$. For the 2×3 case these are the only sets that overestimate and underestimate the given true MeD-set.

Next, let us define two measures which use the notions of overestimation and underestimation. The first measure, *UND*, represents the probability of underestimating the true MeD-set by a procedure. It is defined as

$$UND = Probability\{M \text{ is underestimated}\}.$$

Similarly, the second measure, *OV*, is defined as

$$OV = Probability\{M \text{ is overestimated}\}.$$

Both measures can be obtained by simulation. A proportion of replications when a procedure results in underestimating the true MeD-set, *i.e.*, it returns an estimated MeD-set which overestimates the true one, corresponds to the UND-measure. A proportion of cases when a procedure returns a set underestimating the true one corresponds to the estimated OV-measure.

We use the estimated UND-measure and OV-measure in addition to the estimated FWE, Power, LOP and loss functions to assess the performance of the proposed procedures in our simulation studies presented in Chapter 6.0.

6.0 SIMULATION STUDIES

6.1 GOALS OF THE SIMULATION STUDIES

Since our procedure is based on computing a number of the AVE test statistics, the power of the procedure depends on the power of each single test statistic. As it is shown in Hung *et al.* (1993) (for rectangular designs) and in Appendix D (for non-rectangular designs), the power function of a single AVE test $P\left(T_A > c \middle| \hat{\delta}, \hat{\theta}\right)$ depends on the parameters θ_{ij} only through θ_{AVE} and on a nuisance parameter $\hat{\delta} = (\delta_{11}, \delta_{12}, ..., \delta_{KN})$, where $\delta_{ij} = (\mu_{i0} - \mu_{0j})/\sigma$ and $\delta_{ij} \in (-\infty, \infty)$.

To facilitate the understanding of our simulations we note two results, which are stated by Hung *et al.* (1993) about the power function of a single AVE test for a rectangular design:

Result 6.1. For any given $\hat{\delta}$ with $\delta_{ij} \in (-\infty, \infty)$ and any fixed number $c \in (-\infty, \infty)$, power increases in θ_{AVE}.

Result 6.2. For any given θ_{AVE} and c, the power function increases in each $\delta_{ij} \in (0, \infty)$ and decreases in each $\delta_{ij} \in (-\infty, 0)$ when the remaining δ_{lm} are fixed. Hence, power increases in each $\left|\delta_{ij}\right|$ when the remaining δ_{lm} are fixed.

Result 6.2 does not have any practical value due to the fact that $\hat{\delta}$ is a vector of the control mean differences. In the designs with $K > 1$ and $N > 1$, it is impossible to change just a single component of the vector $\hat{\delta}$ and keep remaining components fixed. So we illustrate the fact that power depends on the values of the control mean differences by changing several components of the vector $\hat{\delta}$ while keeping remaining components fixed.

Our procedure involves simultaneous calculations of several AVE test statistics, and overall power depends on conditional rejection and acceptance of the hypotheses. For example,

in the 2×2 case when the true MeD-set is $\{(2,2)\}$ in order to estimate the set correctly, it is not enough to reject the first tested hypothesis; the procedure must also accept the second tested hypothesis.

The first goal of the simulation study is to confirm an analog of Result 6.1 in terms of the overall power, *i.e.*, to illustrate that the overall power increases in each gain average,

The second goal of the simulation study is related to a special case where our procedure is based on a single AVE test. When the true MeD-set is empty and the first tested hypothesis is accepted, then the procedure correctly estimates the MeD-set as an empty one. In the $K \times N$ case it involves calculating a single test statistic

$$T_A = \frac{\sum_{i.j=1,2} \hat{\theta}_{ij}}{KN\hat{\sigma}}.$$

Since $Power = P(\hat{M} = M)$, in the case when the true MeD-set is empty, the power can be viewed as the probability of accepting the true null hypothesis. Hence, power defined in such a way decreases with increasing $|\delta_{ij}|$'s when the remaining δ_{lm} are fixed. And the second goal of the simulation study is to illustrate this fact.

Table 16. Cases corresponding to familywise error (FWE), lack of power (LOP) and power in the 2×2 case

M \ $\hat{M}$	Empty	{(2,2)}	{(2,1)}	{(1,2)}	{(2,1), (1,2)}	{(1,1)}	Type A
{(2,2)}	LOP	Power	FWE	FWE	FWE	FWE	FWE
{(2,1)}	LOP	LOP	Power	FWE	FWE	FWE	LOP
{(1,2)}	LOP	LOP	FWE	Power	FWE	FWE	LOP
{(2,1), (1,2)}	LOP	LOP	LOP	LOP	Power	FWE	LOP
{(1,1)}	LOP	LOP	LOP	LOP	LOP	Power	LOP

The third goal of the simulation studies is to describe the performance of the proposed procedures based on the regular and modified MPG principle. For this purpose the summary

statistics presented in Chapter 5.0 are computed. Table 16 illustrates which proportions should be included in the estimated FWE, Power, Lack of Power and Type A ambiguity for the 2×2 case. Although, these three summary statistics add up to 1, we report the results in terms of all of these measures. We do not present the table for the 2×3 case, because it is constructed in a similar manner as for the 2×2 case.

In addition to these measures two loss functions, L_1 and L_3, are computed, which are defined in Section 5.3. To specify the values of the misclassification errors for the loss function L_3, we assume the case when it is more desirable to overestimate the true MeD-set than to underestimate it. The weights are set to be $5w_o = w_u$.

Since a sum of all errors corresponding to L_1 is the fixed number for each design, the error values of L_3 are rescaled so that they also add up to this value. For example, Table 14 provides the error values of the first loss function for the 2×2 design. The sum of all errors equals 118, so the errors of L_3 are rescaled so that their total sum is also 118.

In order to simulate the values of L_1 and L_3 for each given true MeD-set, the proportion of replications corresponding to each estimated MeD-set is multiplied by its misclassification error and then added up to produce a single value. The estimates of the loss functions obtained in such a way are the analog of bias defined in a single drug case.

In addition to all of the above measures, the UND-measure and the OV-measure, presented in Section 5.4 are simulated.

The design of the simulation studies and their results are discussed separately for the 2×2 and 2×3 case. In each case the proposed step-down procedure (based on the regular MPG principle) is used for different configurations of the true mean response matrix. In the 2×3 case, a testing procedure based on the modified MPG principle is also considered. In each case the relationship between the overall power and the average gain is examined and the special case of an empty MeD-set is considered. The performance of the procedures is assessed using the simulated measures discussed above. In the 2×3 case the proposed regular and modified step-down procedures are compared. To check the validity of the program code the results of the simulation studies are compared with the numeric computations given by Hung *et al.* (1993) for some specific configurations of the population parameters.

6.2 DESIGN OF THE SIMULATION STUDIES

Throughout the study, the significance level α is fixed at 0.05, $\sigma = 1$ and the common sample size, $n = 30$, is assumed per group. The critical values used in the studies are the critical values given in Table 5 and Sections 4.3 and 4.4, divided by $\sqrt{30}$.

To illustrate the relationship between the overall power and gain averages, we consider $\widehat{\delta} = 0$, where $\delta_{ij} = \mu_{i0} - \mu_{0j}$, so that only settings involving $\mu_{i0} = \mu_{0j}$ with $i, j = 1,2$ for the 2×2 case and $i = 1,2$, $j = 1,2,3$ for the 2×3 case are considered. Without loss of generality, all single drug mean responses, μ_{i0} and μ_{0j}, are taken to be zero. For the special case of an empty MeD-set other values of $\widehat{\delta}$ are considered, which are specified later on.

The single simulation run is performed as follows. For each specified value of the global gain average $\theta_{AVE} = \theta^4_{AVE} = \sum_{i,j=1}^{2} \theta_{ij} / 4$ ($\theta_{AVE} = \theta^6_{AVE} = \sum_{i=1}^{2} \sum_{j=1}^{3} \theta_{ij} / 6$) for the 2×2 (2×3) case, all possible configurations of the MeD-sets are considered. These configurations are given in Sections 4.3 and 4.4.

In the 2×2 case among possible values of the global gain average we consider $\theta_{AVE} = 0,\ 0.4,\ 0.8,\ 1.2$. Since $\theta_{AVE} = 0$ corresponds to only a single MeD-set (empty one) and each of positive $\theta_{AVE} = 0.4,\ 0.8,\ 1.2$ corresponds to five different MeD-sets, there are 16 simulation configurations in total in the 2×2 case. These configurations are presented in Table 17.

In the 2×3 case we consider $\theta_{AVE} = 0,\ 0.4,\ 1.2$. Since there are nine different MeD-sets can be constructed for each of the positive gain average, there are 19 simulation configurations in total.

The performance of the procedure depends not only on the true global average θ_{AVE} but also on all partial averages. For example, in the 2×2 case the partial gain averages are given by $\theta^3_{AVE} = (\theta_{11} + \theta_{12} + \theta_{21}) / 3$, $\theta^{2''}_{AVE} = (\theta_{11} + \theta_{12}) / 2$, $\theta^{2''}_{AVE} = (\theta_{11} + \theta_{21}) / 2$ and $\theta^1_{AVE} = \theta_{11}$. Ideally, it would be desirable to see how the performance of the procedure depends on a single gain average when the rest of the averages are fixed. But changing the value of one gain average leads to changing the value of at least one other average, so it is impossible to construct such a study.

Next, we explain how the population mean dose-response matrix is constructed. For each specified value of the global gain average θ_{AVE} we construct the population mean dose-response matrix such that the corresponding gain matrix has equal positive values for efficacious doses and zero-values for no-gain doses. Such assignment of the mean dose-responses does not differentiate between the cases where the true MeD-sets are given by $\{(1,2)\}$ and $\{(2,1)\}$ (and in addition by $\{(1,3)\}$ and $\{(2,2)\}$ in the 2×3 case). So the results should be similar for these configurations.

A matrix, constructed in such a way, has an isotonic structure, but this specific design should not necessarily improve the results, because the performance of our procedure does not depend on any particular structure of the matrix other than on $\hat{\delta}$ and the gain averages.

Example 6.1. To illustrate how such mean dose-response matrix can be constructed for some non-zero $\hat{\delta} = (\delta_{11}, \delta_{12}, \delta_{21}, \delta_{22})$ with $\delta_{ij} = \mu_{i0} - \mu_{0j}$, consider $\hat{\delta} = (0,0,1,1)$ where $\mu_{10} = 0$, $\mu_{20} = 1$, $\mu_{01} = 0$, and $\mu_{02} = 0$. Suppose, that the global gain average is $\theta_{AVE} = 0.4$ and consider two MeD-sets $M_1 = \{(2,2)\}$ and $M_2 = \{(1,2)\}$. Then, in the 2×2 case the mean dose-response matrices are given by

$$\begin{pmatrix} \times & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 2.6 \end{pmatrix} \text{ and } \begin{pmatrix} \times & 0 & 0 \\ 0 & 0 & 0.8 \\ 1 & 1 & 1.8 \end{pmatrix},$$

where "×" indicates that this dose is not evaluated. The corresponding gain matrices are

$$\Theta_1 = \begin{pmatrix} 0 & 0 \\ 0 & 1.6 \end{pmatrix} \text{ and } \Theta_2 = \begin{pmatrix} 0 & 0.8 \\ 0 & 0.8 \end{pmatrix}.$$

Note, that although, $\theta_{AVE} = \theta^4_{AVE} = 0$ in both cases, the partial gain averages are different: $\theta^3_{AVE} = \theta^{2'}_{AVE} = \theta^{2''}_{AVE} = \theta^1_{AVE} = 0$ in the first case, and $\theta^3_{AVE} = 0.27$, $\theta^{2'}_{AVE} = 0.4$, $\theta^{2''}_{AVE} = \theta^1_{AVE} = 0$ in the second case . This concludes the example.

Next, for the specified $\mu_{ij}'s$ the group sample means

$$\overline{Y} = \begin{pmatrix} \times & \bar{y}_{01} & \bar{y}_{02} \\ \bar{y}_{10} & \bar{y}_{11} & \bar{y}_{12} \\ \bar{y}_{20} & \bar{y}_{21} & \bar{y}_{22} \end{pmatrix} \text{ or } \overline{Y} = \begin{pmatrix} \times & \bar{y}_{01} & \bar{y}_{02} & \bar{y}_{03} \\ \bar{y}_{10} & \bar{y}_{11} & \bar{y}_{12} & \bar{y}_{13} \\ \bar{y}_{20} & \bar{y}_{21} & \bar{y}_{22} & \bar{y}_{23} \end{pmatrix}$$

are generated depending on the design, where the $\bar{y}_{ij}$ are independent random variables such that $\bar{y}_{ij} \sim N(\mu_{ij},\ 1/\sqrt{30})$. Then, the step-down procedure is applied to these data. In the 2×3 case, both proposed procedures, regular and modified, are applied to these same data. The result of each procedure is noted and the above steps are replicated 100,000 times for the specified value of $\hat{\delta}$ and θ_{AVE}.

To illustrate the fact that power depends on the values of the control mean differences, in addition to the simulation studies described above, we consider a special case when the true MeD-set is empty ($\theta_{AVE} = 0$) and so that the correct estimation involves only a single step procedure. In the 2×2 case we consider this situation for two configurations of the nuisance parameter $\hat{\delta}$, $\hat{\delta} = (0,0,1,1)$ and $\hat{\delta} = (1,1,1,1)$. The results corresponding to these configurations are then combined with the results of the first simulation study where $\theta_{AVE} = 0$ and $\hat{\delta} = 0$, and are analyzed together. Table 17 presents all configurations of the population parameters used in the 2×2 case.

In order to check the validity of the program, in the 2×3 case we consider four additional cases with non-zero nuisance parameter $\hat{\delta}$ given by Hung *et al.* (1993). These authors calculate the exact power for the case of different global gain averages and specified values of the control mean differences. We consider just some of these cases given in Table 18.

These special cases are discussed in more details when the results of the corresponding simulation runs are presented.

For all the combination drug studies 100,000 simulations are used for estimating the measures involving probabilities, this leads to a maximum standard error of estimation of $\sqrt{0.5(0.5)/100{,}000} = 0.002$. Since the loss functions, L_1 and L_3, are not based on probabilities, no standard errors are herein provided.

Table 17. Population parameters configurations used in the simulation studies in the 2×2 case

MeD-Set	θ_{AVE}	$\widehat{\delta}$
Empty	0	$\widehat{\delta} = (0,0,0,0)$ $\widehat{\delta} = (0,0,1,1)$ $\widehat{\delta} = (1,1,1,1)$
{(1,1)}	0.4 0.8 1.2	$\widehat{\delta} = (0,0,0,0)$
{(1,2)} ({(2,1)})	0.4 0.8 1.2	$\widehat{\delta} = (0,0,0,0)$
{(2,1), (1,2)}	0.4 0.8 1.2	$\widehat{\delta} = (0,0,0,0)$
{(2,2)}	0.4 0.8 1.20	$\widehat{\delta} = (0,0,0,0)$

Table 18. Additional population parameters configurations used in the simulation studies in the 2×2 case

θ_{AVE}	μ_{10}, μ_{20}	μ_{01}, μ_{02}, μ_{03}
$\theta_{AVE} = 0$	0.1, 1.0	0.1, 0.3, 1.0
$\theta_{AVE} = 0$	0.1, 0.3	0.1, 0.2, 0.3
$\theta_{AVE} = 0.2$	0.1, 0.3	0.1, 0.2, 0.3
$\theta_{AVE} = 0.6$	0.1, 0.3	0.1, 0.2, 0.3

6.3 DETAILED RESULTS OF THE 2 BY 2 SIMULATION STUDIES

First, we discuss the simulation results regarding the relationship between the value of the gain average and the power of the procedure. Table 19, Table 20 and Table 21 summarize the results of the studies. Note, that in the cases where the true MeD-sets are given by $\{(1,2)\}$ and $\{(2,1)\}$ the procedure produces very similar results, as expected.

Table 19 contains the sample FWE (upper entry) and sample power (lower entry) for each configuration of the gain average and the population MeD-set. Since all FWE's are less than $0.05 + 1.96\sqrt{0.05 \cdot 0.95 / 100{,}000} = 0.051$ the procedure again demonstrates the theory that the FWE is strongly controlled at 0.05. Since configurations with MeD-set given by $\{(1,1)\}$ involve no FWE, Table 19 has no upper entry for such configurations.

The power of the procedure greatly depends on the value of θ_{AVE}. If $\theta_{AVE} = 0$ then the procedure correctly estimates the empty MeD-set with a high probability.

The proposed procedure has the average power of 0.366 for $\theta_{AVE} = 0.4$, 0.833 for $\theta_{AVE} = 0.8$ and 0.983 for $\theta_{AVE} = 1.2$. So for all non-zero effect averages, the power of the procedure increases with the value of the gain average.

Among the considered designs, our procedure identifies the true MeD-set $\{(1,1)\}$, $\{(1,2)\}$ ($\{(2,1)\}$), $\{(2,1),\ (1,2)\}$ and $\{(2,2)\}$ with the average power of 0.675, 0.671, 0.485 and 0.866, respectively. Hence, among the non-empty sets, the MeD-set given by $\{(2,2)\}$ is correctly identified with the highest probability. The procedure performs the worst in the case when the true MeD-set is "diagonal", *i.e.*, is given by $\{(2,1),\ (1,2)\}$.

Table 20 contains the sample probabilities of Type A ambiguity (upper entry) and LOP (lower entry). Probability of Type A outcome decreases in θ_{AVE} and it is very small, 0.001, when the MeD-set is empty. The average probability of Type A outcome is 0.068 for $\theta_{AVE} = 0.4$, 0.042 for $\theta_{AVE} = 0.8$ and 0.002 for $\theta_{AVE} = 1.2$. Configurations with MeD-set $\{(1,1)\}$, $\{(1,2)\}$ ($\{(2,1)\}$), $\{(2,1),\ (1,2)\}$ and $\{(2,2)\}$ have the average probability of Type A outcome of 0.028, 0.009, 0.140 and 0.002, respectively. So the configurations with true "diagonal" MeD-set correspond to the greatest frequency of Type A ambiguity.

As expected, similar conclusions can be drawn based on the estimated lack of power. The proposed step-down procedure has the average lack of power of 0.630 ($\theta_{AVE} = 0.4$), 0.162 ($\theta_{AVE} = 0.8$) and 0.013 ($\theta_{AVE} = 1.2$). Independent of the gain average, cases where $M = \{(2,2)\}$ correspond to the smallest value of the lack of power. Configurations with the "diagonal" MeD-set, $M = \{(2,1),\ (1,2)\}$, correspond to the greatest value of the lack of power.

Table 21 contains the estimated values of the loss functions, discussed in Section 5.3. Two loss functions behave similarly: their values decrease with increasing gain average. Moreover, if one function increases (decreases) then the other also does. The procedure has the average losses values of $L_1 = 2.727$, $L_3 = 0.813$ (when $\theta_{AVE} = 0.4$); $L_1 = 0.287$, $L_3 = 0.119$ (when $\theta_{AVE} = 0.8$); and $L_1 = 0.035$, $L_3 = 0.027$ (when $\theta_{AVE} = 1.2$). Again, independent of the value of the gain average, the cases where the MeD-set is $M = \{(2,1),\ (1,2)\}$ correspond to the largest loss among possible configurations of the MeD-sets. For all considered non-zero gain averages, the true $M = \{(2,2)\}$ is identified with the smallest loss.

Finally, let us mention the results of additional simulation studies, summarized in Table 22. When there is no MeD ($\theta_{AVE} = 0$) there is a decrease in "power" of the procedure associated with the increased values of components of $\widehat{\delta}$.

Summary of the 2×2 Simulation Results

1) The estimation quality of the procedure increases with the increased positive gain average. If the non-zero gain averages are relatively large then the procedure correctly identifies the MeD-set with high power.
2) Among possible configurations in the 2×2 case, the empty MeD-set and the set containing the highest combination, $\{(2,2)\}$, are correctly identified with the highest probability.
3) The "diagonal" configuration of the MeD-set, $M = \{(2,1),\ (1,2)\}$, is the most difficult to be estimated by the procedure.
4) Additional simulation studies indicate that the results of the estimation depend on the values of the mean dose-response differences of the single drugs.

Table 19. Estimated familywise error (the upper entry) and power (the lower entry) for $\tilde{\delta} = 0$

θ_{AVE}	MeD-set	Step-Down
0	Empty	Not Defined 0.996
0.4	{(1,1)}	 0.170
	{(1,2)}	0.004 0.191
	{(2,1)}	0.004 0.189
	{(2,1), (1,2)}	0.009 0.044
	{(2,2)}	0.005 0.608
0.8	{(1,1)}	 0.859
	{(1,2)}	0.004 0.831
	{(2,1)}	0.004 0.828
	{(2,1), (1,2)}	0.012 0.488
	{(2,2)}	0.005 0.995
1.2	{(1,1)}	 0.997
	{(1,2)}	0.004 0.992
	{(2,1)}	0.004 0.992
	{(2,1), (1,2)}	0.012 0.923
	{(2,2)}	0.004 0.996

Table 20. Proportion of cases resulted in Type A ambiguity (upper entry) and lack of power (lower entry)

θ_{AVE}	MeD-set	Step-Down
0	Empty	0.001 Not Defined
0.4	{(1,1)}	0.078 0.830
	{(1,2)}	0.024 0.805
	{(2,1)}	0.023 0.807
	{(2,1), (1,2)}	0.213 0.947
	{(2,2)}	0.002 0.388
0.8	{(1,1)}	0.006 0.141
	{(1,2)}	0.002 0.165
	{(2,1)}	0.002 0.168
	{(2,1), (1,2)}	0.199 0.500
	{(2,2)}	0.002 0.001
1.2	{(1,1)}	0.000 0.003
	{(1,2)}	0.000 0.004
	{(2,1)}	0.000 0.004
	{(2,1), (1,2)}	0.009 0.065
	{(2,2)}	0.002 0.000

Table 21. Estimated values of the loss functions based on Definition 5.1 (upper entry) and Definition 5.3 (lower entry) of Section 5.3

θ_{AVE}	MeD-set	Step-Up	Step-Down
0	Empty	0.126 0.204	0.018 0.020
0.4	{(1,1)}	3.299 1.334	4.097 1.651
	{(1,2)}	2.719 0.783	3.124 0.842
	{(2,1)}	2.726 0.786	3.134 0.845
	{(2,1), (1,2)}	4.304 1.214	4.433 1.236
	{(2,2)}	1.607 0.415	1.558 0.286
0.8	{(1,1)}	0.326 0.120	0.356 0.141
	{(1,2)}	0.100 0.119	0.338 0.135
	{(2,1)}	0.098 0.112	0.342 0.138
	{(2,1), (1,2)}	0.660 0.267	0.660 0.267
	{(2,2)}	0.060 0.142	0.010 0.013
1.2	{(1,1)}	0.008 0.002	0.008 0.002
	{(1,2)}	0.028 0.092	0.016 0.023
	{(2,1)}	0.029 0.093	0.015 0.024
	{(2,1), (1,2)}	0.149 0.082	0.149 0.082
	{(2,2)}	0.054 0.135	0.007 0.013

Table 22. Estimated power of the procedure in the case of $\theta_{AVE} = 0$ for different values of the mean control differences

$\hat{\delta}$	Estimated Power of the Step-Down Procedure
$\hat{\delta} = (0,0,0,0)$	0.996
$\hat{\delta} = (0,0,1,1)$	0.986
$\hat{\delta} = (1,1,1,1)$	0.949

6.4 DETAILED RESULTS OF THE 2 BY 3 SIMULATION STUDIES

Tables 23-25 summarize the results of the simulation studies for the 2×3 case.

First, let us compare the performance of the regular and modified step-down procedures. The modified procedure uses the generalized MPG testing principle presented in Section 4.6. The only difference between the procedures is that the regular MPG procedure may result in Type B ambiguity (6^{th} or 9^{th} "decision" sequences of Table 7) but the modified procedure provides meaningful estimates of the MeD-set ($\{(1,3)\}$ or $\{(2,1)\}$, respectively) in such cases.

Since the probability of Type B ambiguity outcome is very small for $\theta_{AVE} = 0.8$ and $\theta_{AVE} = 1.2$, the simulated measures corresponding to the modified procedure are given for $\theta_{AVE} = 0$ and $\theta_{AVE} = 0.4$ only. These are presented in Table 23. Since the occurrence of Type A ambiguity does not change whether or not the regular or modified MPG principle is used, Table 23 presents a single entry of simulated probability of Type A ambiguity for each population configuration.

The modified procedure strongly controls the FWE at $\alpha = 0.05$. The performance of this procedure is very similar to the one of the regular procedure in terms of estimated values of power, FWE and lack of power. When the true MeD-set is given by $\{(1,3)\}$ or $\{(2,1)\}$ the power (lack of power) of the modified procedure is greater (less) than the power (lack of power) of the regular procedure. On the other hand, when the true MeD-set is given by $\{(1,3)\}$ or $\{(2,1)\}$, Type B ambiguity occurs in less than 1% of the replications so the difference in power for the regular and modified procedures is not significant. For example, when the true MeD-set is given by $\{(1,3)\}$, the power is 0.151 and 0.157 for the regular and modified procedures, respectively.

Table 23. Simulation results of the regular and modified procedures for $\theta_{AVE} = 0$ and $\theta_{AVE} = 0.4$ in the 2×3 case

Population MeD-Set	Empty	(2,3)	(2,2)	(1,3)	(2,1)	(2,2) (1,3)	(1,2)	(2,1) (1,3)	(2,1) (1,2)	(1,1)
Power	0.999	0.670	0.165	0.151	0.075	0.041	0.079	0.022	0.025	0.147
$Power_m$	0.998	0.670	0.165	0.157	0.081	0.042	0.078	0.022	0.025	0.146
LOP	0.000	0.328	0.833	0.845	0.924	0.955	0.918	0.975	0.968	0.853
LOP_m	0.000	0.328	0.833	0.839	0.918	0.954	0.918	0.975	0.967	0.854
FWE	0.000	0.002	0.002	0.004	0.001	0.004	0.004	0.004	0.008	0.000
FWE_m	0.000	0.002	0.002	0.004	0.001	0.005	0.003	0.004	0.008	0.000
L_1	0.007	1.317	2.969	2.972	4.384	3.998	3.922	4.614	5.276	4.750
$L_{1,m}$	0.009	1.317	2.968	2.976	4.369	4.004	3.944	4.798	5.648	4.881
L_3	0.004	0.102	0.350	0.355	0.564	0.501	0.681	0.645	0.795	0.988
$L_{3,m}$	0.005	0.103	0.350	0.356	0.562	0.503	0.696	0.679	0.886	1.043
UND	0.001	0.001	0.001	0.001	0.000	0.002	0.002	0.004	0.008	0.000
UND_m	0.001	0.002	0.001	0.001	0.000	0.003	0.002	0.004	0.008	0.000
OV	0.000	0.328	0.827	0.826	0.911	0.616	0.754	0.509	0.572	0.741
OV_m	0.000	0.328	0.828	0.827	0.910	0.619	0.758	0.510	0.572	0.797
Type A	0.000	0.000	0.006	0.012	0.008	0.250	0.153	0.233	0.157	0.056
Type B	0.000	0.000	0.000	0.008	0.005	0.006	0.011	0.063	0.131	0.056

Table 24. Simulation results of the proposed procedure for $\theta_{AVE} = 0.8$ in the 2×3 case

Population MeD-Set	(2,3)	(2,2)	(1,3)	(2,1)	(2,2) (1,3)	(1,2)	(2,1) (1,3)	(2,1) (1,2)	(1,1)
Power	0.998	0.832	0.826	0.553	0.466	0.526	0.317	0.364	0.859
LOP	0.000	0.166	0.170	0.446	0.529	0.470	0.679	0.624	0.141
FWE	0.002	0.002	0.004	0.001	0.005	0.004	0.004	0.012	0.000
L_1	0.005	0.334	0.343	0.920	0.444	1.260	0.891	0.733	0.389
L_3	0.003	0.052	0.059	0.142	0.073	0.157	0.111	0.133	0.065
UND	0.001	0.001	0.001	0.000	0.003	0.002	0.004	0.012	0.000
OV	0.000	0.165	0.167	0.444	0.014	0.434	0.056	0.154	0.135
Type A	0.000	0.001	0.003	0.002	0.319	0.036	0.361	0.237	0.005
Type B	0.000	0.000	0.001	0.000	0.002	0.000	0.020	0.061	0.002

Table 25. Simulation results for $\theta_{AVE} = 1.2$ in the 2×3 case

Population MeD-Set	(2,3)	(2,2)	(1,3)	(2,1)	(2,2) (1,3)	(1,2)	(2,1) (1,3)	(2,1) (1,2)	(1,1)
Power	0.998	0.996	0.994	0.937	0.916	0.914	0.805	0.847	0.997
LOP	0.000	0.003	0.003	0.062	0.079	0.082	0.191	0.141	0.003
FWE	0.002	0.002	0.003	0.001	0.004	0.004	0.004	0.012	0.000
L_1	0.004	0.008	0.012	0.126	0.114	0.250	0.387	0.262	0.008
L_3	0.003	0.003	0.008	0.022	0.021	0.033	0.048	0.052	0.001
UND	0.001	0.001	0.001	0.000	0.003	0.002	0.004	0.012	0.000
OV	0.000	0.003	0.003	0.062	0.000	0.081	0.022	0.054	0.003
Type A	0.000	0.000	0.001	0.000	0.028	0.001	0.059	0.027	0.000
Type B	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000

Table 26. Estimated average loss

Average Loss	$\theta_{AVE} = 0$	$\theta_{AVE} = 0.4$	$\theta_{AVE} = 0.8$	$\theta_{AVE} = 1.2$
$\overline{L}_1$	0.007	3.800	0.591	0.130
$\overline{L}_{1,m}$	0.009	3.878	0.615	0.129
$\overline{L}_3$	0.004	0.553	0.088	0.021
$\overline{L}_{3,m}$	0.005	0.575	0.094	0.022

Type B ambiguity happens the most frequently when the true MeD-set is given by the "diagonal" set $\{(1,2),\ (2,1)\}$. When $\theta_{AVE} = 0.4$ the probability of Type B outcome is about 13%, this percentage is a sum of two percentages. The first percentage corresponds to the cases when the regular procedure results in the 6th "decision" sequence of Table 7 (about 3%) and the second percentage corresponds to the cases when it results in the 9th "decision" sequence (10%). The estimated probability of Type B ambiguity is relatively high for the MeD-set $\{(1,2),\ (2,1)\}$ due to two reasons. The first reason is that the diagonal designs require the testing of more hypotheses. The second reason is the small value of the gain average. Indeed, when the $\theta_{AVE} = 0.8$, the percentage of Type B outcomes reduces to 6% and when $\theta_{AVE} = 1.2$ it reduces to 0.1%.

The power of each procedure depends on the value of the gain average. In the case of a zero gain average, the estimated power is greater than 0.99. The estimated values of the average power are 0.246, 0.638 and 0.934 for $\theta_{AVE} = 0.4$, $\theta_{AVE} = 0.8$ and $\theta_{AVE} = 1.2$, respectively.

The estimated averages of the loss functions are summarized in Table 26, where $\overline{L}_i$, $i = 1,3$ denotes the average loss corresponding to a regular procedure and $\overline{L}_{i,m}$, $i = 1,3$ denotes the average loss of the modified procedure.

The loss averages seem to be higher for the modified procedure than for the regular one in the case of $\theta_{AVE} = 0.4$. These higher values are due to the increased proportion of cases corresponding to the meaningful MeD-set estimates.

In terms of underestimating the true MeD-set, both procedures perform similarly. There are technical reasons for why this happens. The only true sets that can be underestimated by the set $\{(1,3)\}$ are the set $\{(2,3)\}$ and an empty set. But Type B outcome almost never happens for such configurations (proportion of Type B outcome is less than 0.1%). Similarly, the only true sets that are underestimated by the set $\{(2,1)\}$ are the sets $\{(2,2)\}$ and $\{(2,3)\}$. Again, Type B ambiguity happens in less than 0.1% of the cases for such configurations.

In terms of overestimating the true MeD-sets, the results of the procedures are slightly different when the true MeD-set is given by $\{(1,1)\}$. In such cases $\{(1,1)\}$ is overestimated by both sets, $\{(1,3)\}$ and $\{(2,1)\}$, and type B outcome happens about 6% of the time, so the

modified procedure results in higher values of the estimated $OV-measure$ than the regular procedure does.

Tables 23-25 also contain the frequencies of Type A ambiguity. When there is no MeD, the proposed procedure almost never results in such outcome (the proportion of such cases is less than 0.1%). The average frequency of Type A ambiguity is 0.097 for $\theta_{AVE}=0.4$, 0.107 for $\theta_{AVE}=0.8$ and 0.013 for $\theta_{AVE}=1.2$. Among possible configurations, the configurations with the true "diagonal" MeD-sets, $\{(2,2),\ (1,3)\}$, $\{(2,1),\ (1,2)\}$ or $\{(2,1),\ (1,3)\}$, correspond to the highest average frequency of Type A ambiguity. Among those, configurations with the true MeD-set $\{(2,1),\ (1,3)\}$ correspond to the highest probability of Type A outcome. The configurations with the true MeD-sets given by $\{(2,3)\}$, $\{(2,2)\}$, $\{(2,1)\}$ and $\{(1,3)\}$ correspond to the lowest average frequency of Type A ambiguity, less than 1%, followed by the set $\{(1,1)\}$ (2%) and $\{(1,2)\}$ (5%).

These conclusions are summarized at the end of this section.

Table 27. Simulation probability and the exact power of a single AVE test statistic given by Hung *et al.* (1993)

θ_{AVE}	$\mu_{10},\ \mu_{20}$	$\mu_{01},\ \mu_{02},\ \mu_{03}$	Simulation Probability	Power of a Single AVE Test
$\theta_{AVE}=0$	0.1, 1.0	0.1, 0.3, 1.0	0.990	0.01
$\theta_{AVE}=0$	0.1, 0.3	0.1, 0.2, 0.3	0.996	0.005
$\theta_{AVE}=0.2$	0.1, 0.3	0.1, 0.2, 0.3	0.822	0.18
$\theta_{AVE}=0.6$	0.1, 0.3	0.1, 0.2, 0.3	0.009	0.99

To confirm the correctness of our simulation program we consider four special cases given in Table 18. Table 27 presents the simulation probabilities for such configurations and the exact power obtained by Hung *et al.* (1993). In the first two cases with $\theta_{AVE}=0$ the simulated probabilities provide the estimated "power" of the procedure, both simulated values agree with Hung's exact power of a single AVE test, *i.e.*, $Power=P(\hat{M}=\varnothing \mid M=\varnothing)=1-Power_{HUNG}$.

Based on the simulated probabilities for the last two cases we can conclude that the first tested hypothesis is rejected with estimated probabilities of $1 - 0.822 = 0.178$ and $1 - 0.009 = 0.991$. Again, both values agree with Hung's values, 0.18 and 0.99.

6.5 SUMMARY OF THE RESULTS

We summarize the conclusions presented in Section 6.3 for the 2×2 case and Section 6.4 for the 2×3 case as follows.

1) The quality of the proposed procedures improves with increasing positive gain average.
2) Our proposed procedures perform the best in the case when there is no MeD.
3) Among possible non-empty true MeD-sets, the set containing the highest combination is correctly identified with the highest probability. The "diagonal" configurations of the MeD-set are correctly identified with the lowest probability.
4) The regular and modified step-down procedures perform very similarly. The modified procedure results in a slightly greater average loss but provides a meaningful MeD-set estimate in the cases when the regular procedure results in ambiguities.
5) While the theory demonstrates that our procedure controls the FWE rate, the simulations reinforce this.
6) The probability of Type A ambiguity is small when the gain averages are relatively large. The population configurations with "diagonal" MeD-set correspond to the highest frequency of Type A ambiguity. Configurations with true MeD-set containing the highest combination correspond to the smallest probability of Type A ambiguity.
7) Additional simulation studies indicate that the results of the procedures depend on the values of the mean dose-response differences of the single drugs.
8) The program validation is checked using the exact power values presented by Hung *et al.* (1993).

7.0 DISCUSSION

7.1 DISCUSSION OF THE RESULTS

In the biopharmaceutical industry in order for a new drug to be marketed, efficacy and safety studies are required. One of the main goals of the efficacy study is to identify the effective dose range, *i.e.*, all doses which are superior to placebo. This applies to both a single drug and combination drug setting.

In addition to identifying all effective doses, clinicians are interested in selecting the minimum effective dose. Such a dose is especially desirable when increasing the dose level leads to the serious side effects. In a single drug case, the minimum effective dose, if it exists, is unique.

We propose a closed testing procedure for identifying the minimum effective dose in a single drug study. Our procedure uses the assumption of non-decreasing mean dose-response, which is the most popular shape of the mean dose-response function. The proposed procedure is based on the closed testing principle applied to a family of hypotheses and involves constructing simultaneous confidence bands. Our procedure performs better than its competitors when the population mean dose-response function is rapidly rising near zero, *i.e.*, when the true minimum effective dose is relatively low, which is often the case in biopharmaceutical trials.

In the case of a combination drug, in addition to demonstrating safety and efficacy the FDA requires demonstrating that each component makes a contribution to the claimed effects. A combination which satisfies the last requirement is called an efficacious combination. The efficacious combinations with the lowest doses are called the minimum efficacious combinations. Since the minimum efficacious combination may not be unique, we introduce the notion of the minimum efficacious dose set (MeD-set).

The problem of identifying the MeD-set is very complex. The more dose levels of each drug, there are more different structures of the population MeD-sets that will need to be identified by a procedure.

We propose a closed testing procedure for identifying the MeD-set. Our procedure requires first, identifying all possible patterns of the population MeD-set for a given design. Next, the closed hypothesis family is constructed and the proper step-down partial testing order is specified. Then, the hypotheses are tested using the closed testing scheme. The pattern of rejected and accepted hypotheses provides the estimated MeD-set.

The main advantage of the procedure is the strong control of overall error at a specified level of significance. This feature is crucial in multiple testing because when the number of hypotheses to be tested is large, the familywise error can rapidly increase. As a consequence, a procedure involving multiple testing results in false conclusion about the drug effectiveness with high probability.

The other advantage of the proposed procedure is that in the case of a combination consisting of approved single drugs, our procedure demonstrates effectiveness and efficacious of the combination drug, so the two regulatory problems stated by the FDA can be solved at once. That is because in the case of approved component drugs, it is known that each single drug is superior to placebo, and therefore, all efficacious combinations are also superior to placebo, *i.e.*, they are effective by implication. So the proposed procedure identifies the minimum efficacious and therefore, effective combinations.

There are some complexities associated with the hypothesis testing approach used to identify the MeD-set. First, when the number of active doses of each drug is high, there are a large number of hypotheses to be tested. While in the case of four combinations (two active doses of one single drug and two active doses of the other single drug) there are five hypotheses in the hypothesis family; in the case of nine combinations (three active doses of one single drug and three active doses of the other single drug), there are nineteen hypotheses in the family.

The other technical complexity of our procedure is related to a complicated structure of the parameters for the population MeD-set. The procedure is constructed so that all possible patterns of the parameters can be identified, but together with those estimates it may result in ambiguities. We describe three types of ambiguities that may be present in the designs with up to

nine combinations. One type of these ambiguities can be eliminated by generalizing the testing scheme; others require additional research.

These two complexities, large number of hypotheses to be tested and possible resulting in ambiguity, are of primary concern when the number of combinations is large. But the modern biopharmaceutical combination drug studies usually do not involve more than four or six drug combinations, and for such designs our proposed procedure is quite effective.

Since there are no other procedures for identifying the minimum efficacious doses in the combination drug case, the performance of the proposed procedure can not be compared with the performance of any other procedure. Instead, Monte-Carlo simulations are used in order to describe the properties and goodness of the estimation.

The simulation studies reinforce the applicability of the proposed procedure. Among possible structures of the population MeD-set, the sets containing the highest dose combinations are identified by the procedure the best. Also, if there are no efficacious combinations in the study, the procedure detects this case with high power.

7.2 FUTURE RESEARCH

The primary goal of our future research is to continue developing statistical techniques that can be used in the analysis of the efficacy studies of the combination drug. The problems related to this goal are discussed in this section.

The first goal is to construct a similar step-down testing procedure based on the MAX test statistic. The proposed procedure is based on the closed testing principle, which allows testing a single hypothesis at a fixed level of significance. The testing of a single hypothesis involves constructing the AVE statistic proposed by Hung *et al.* (1993). As it is mentioned in Section 4.1, Hung *et al.* (1993) consider one more test statistic, the MAX statistic, for testing individual hypotheses.

In order to implement the MAX test, we need to obtain the distribution function corresponding to non-rectangular designs; then define the power function; and finally, provide the formula for calculating the critical values for non-rectangular designs. Using Hung's *et al.*

(1993) results for testing the hypotheses corresponding to the rectangular designs, we can test all of the hypotheses under the closed testing principle.

Finally, the two procedures, based on the AVE and MAX test statistics, can be compared by simulation. We plan to obtain the conditions and identify the population patterns, when one procedure dominates the other in terms of the measures presented in Chapter 5.0.

The long term research goal is to develop the confidence bands techniques similar to the ones introduced in the single drug case and presented in Section 2.4. By constructing the simultaneous lower bands for the dose-response surface we could estimate the population mean dose-response function. To construct such bands we can start with the simplest case of monotone mean dose-response surface. Based on the knowledge obtained in this case a general setting can be considered.

In this dissertation we present a testing procedure for identifying all minimum efficacious doses in the combination drug study. This procedure uses an assumption of approved single drugs. So the regulatory problems stated by the FDA, requiring demonstration that the combination is superior to placebo, and that the combination is superior to each component drug taken alone, can be handled by our procedure. That is because in the case of approved component drugs, it is known that each single drug is superior to placebo, and therefore, all efficacious combinations are also superior to placebo. The other long term research goal is to construct a procedure for identifying the minimum efficacious doses, similar to the one proposed in this dissertation, but without the assumption that each component drug is superior to placebo.

APPENDIX A

R CODE FOR SIMULATION THE UPPER CRITICAL VALUES FOR THE TWO-SIDED AND ONE-SIDED SIMULTANEOUS MAX-MIN CONFIDENCE BANDS FOR THE CASE OF EQUAL SAMPLE SIZES AND KNOWN VARIANCE

```
sd1<-(1/(2^(0.5)));
max1<-500000;
k<-6; alp<-0.95;
c1<-as.array(k);
c2<-as.array(k);
t<-as.array(max1);
w<-as.array(max1);
for (r in 1:k){
                    A<-matrix(0,r,r);
                    s<-(r+1);
                    y<-as.array(s);
                    z<-as.array(r);
                    for (l in 1:max1){
                     y<-rnorm(s, mean=0,sd=sd1);
                     for(n in 1:r){
                     z[n]<-(y[n]-y[s]);
                                 }
                     for (b in 1:r) { d<-b; A[b,d]<-z[b];
                                    }
```

```
     for (i in 1:r){
     for (j in i:r){
     if (i<j) {A[i,j]<-((((j-i)^0.5)*(A[i,(j-1)])+z[j])/((j-
i+1)^0.5));
                    A[j,i]<-A[i,j]}
                                  }
                                  }
                    w[l]<-max(A); t[l]<-max(abs(A));
                           }
                    c1[r]<-quantile(w,alp); c2[r]<-
quantile(t,alp);
                           }
     print(c1); print(c2);
```

Comments:

`sd1` specifies the standard deviation of the group sample mean;

`max1` stores the number of replications;

`k` defines the number of active doses;

`alp` specifies the level of significance;

`c1` is a vector of length max1 of the upper one-sided critical values, c1[i] stores the critical value based on i active doses;

`c2` is a vector of length max1 of the upper two-sided critical values, c2[i] stores the critical value based on i active doses.

APPENDIX B

R CODE FOR THE SIMULATION STUDIES FOR THE SINGLE DRUG CASE

```
sd1<-1;
N<-matrix(0,6,15);

for (t in 1:5) {
   N[5,t]<-3;
}
for (t in 6:10) {
   N[5,t]<-4;
}
for (t in 11:15) {
   N[5,t]<-5;
}
N[4,2]<-1.5;
N[3,3]<-1;
N[4,3]<-2;
N[2,4]<-0.75;
N[3,4]<-1.5;
N[4,4]<-9/4;
N[1,5]<-0.6;
N[2,5]<-1.2;
N[3,5]<-1.8;
```

```
N[4,5]<-2.4;
N[4,7]<-2;
N[3,8]<-4/3;
N[4,8]<-8/3;
N[2,9]<-1;
N[3,9]<-2;
N[4,9]<-3;
N[1,10]<-4/5;
N[2,10]<-8/5;
N[3,10]<-12/5;
N[4,10]<-16/5;
N[4,12]<-5/2;
N[3,13]<-5/3;
N[4,13]<-10/3;
N[2,14]<-5/4;
N[3,14]<-10/4;
N[4,14]<-15/4;
N[1,15]<-1;
N[2,15]<-2;
N[3,15]<-3;
N[4,15]<-4;
print(N);
T<-5;
Delta<-3;
POPMEAN<-as.array(6);
lee<-as.array(10);
korn<-as.array(10);
wilm<-as.array(10);
"Critical values for up to 10 active doses are given"
"Only first five are used by the program"
lee[1]<-1.645;
lee[2]<-2.12;
```

```
lee[3]<-2.45;
lee[4]<-2.72;
lee[5]<-2.98;
lee[6]<-3.18;
lee[7]<-3.39;
lee[8]<-3.59;
lee[9]<-3.76;
lee[10]<-3.94;
korn[1]<-1.645;
korn[2]<-1.92;
korn[3]<-2.06;
korn[4]<-2.16;
korn[5]<-2.23;
korn[6]<-2.29;
korn[7]<-2.34;
korn[8]<-2.38;
korn[9]<-2.42;
korn[10]<-2.72;
wilm[1]<-1.645;
wilm[2]<-1.716;
wilm[3]<-1.739;
wilm[4]<-1.750;
wilm[5]<-1.756;
wilm[6]<-1.760;
wilm[7]<-1.763;
wilm[8]<-1.765;
wilm[9]<-1.767;
wilm[10]<-1.768;
y<-as.array(6);
z<-as.array(5);
M<-0;
E<-as.array(1);
```

```
E0<-as.array(1);
E1<-as.array(1);
E2<-as.array(1);
E3<-as.array(1);
E4<-as.array(1);
E5<-as.array(1);
E6<-as.array(1);
G<-as.array(1);
Q<-as.array(1);
t<-as.array(5);
LEST<-as.array(6);
LDFEST<-as.array(6);
KEST<-as.array(6);
WEST<-as.array(6);
X<-matrix(-100,5,5);
A<-matrix(0,5,5);
V<-matrix(0,5,5);
B<-as.array(5);
set.seed(100)
max1<-10000;
for (n in 1:15) {
   if ((n==1)||(n==6)||(n==11)) {T<-5};
   if ((n==2)||(n==7)||(n==12)) {T<-4};
   if ((n==3)||(n==8)||(n==13)) {T<-3};
   if ((n==4)||(n==9)||(n==14)) {T<-2};
   if ((n==5)||(n==10)||(n==15)){T<-1};
   if ((n==6)||(n==7)||(n==8)||(n==9)||(n==10)){Delta<-4};
   if ((n==11)||(n==12)||(n==13)||(n==14)||(n==15)){Delta<-
5};
   for (u in 1:6) {
      POPMEAN[u]<-N[u,n]
   }
```

```
LEXP<-0;
LDFEXP<-0;
KEXP<-0;
WEXP<-0;
LBIAS<-0;
KBIAS<-0;
WBIAS<-0;
LDFBIAS<-0;
for (k in 1:6) {
   WEST[k]<-0;
   LEST[k]<-0;
   KEST[k]<-0;
   LDFEST[k]<-0;
}
LMED<-as.array(max1);
KMED<-as.array(max1);
WMED<-as.array(max1);
LDFMED<-as.array(max1);
E<-0;
E0<-0;
E1<-0;
E2<-0;
E3<-0;
E4<-0;
E5<-0;
E6<-0;
G<-0;
Q<-0;
for (l in 1:max1) {
   for (i in 1:6) {
      y[i]<-rnorm(1,mean=POPMEAN[i],sd=sd1)
   }
```

```
      "LEE PROCEDURE"
#Generating A (needed to calculate maximums)
      for (b in 1:5) {
         A[b,b]<-(y[b]-y[6]);
      }
      for (r in 1:5) {
         for (j in r:5) {
            if (j>r) {h<-(j-r+1);A[r,j]<-((((j-r)*(A[r,j-
1]))+y[j]-y[6])/h)}
            }
      }
      "Step-down Lee"
      p<-1; # to break the while loop
       j<-5;
      while ((j>=1)&(p>0)) {
         M<-(y[1]-y[6]-lee[j]*(2^0.5));
         for (k in 1:j) {
            for (h in 1:k) {
               w<-(k-h+1);
               F<-(A[h,k]-(lee[j])*((2/w)^0.5));
              if (F>M) {M<-F}
              }
         }
         if (M>=0) {j<-(j-1)} else {p<-0}
     }
        LMED[l]<-(j+1);
      h<-LMED[l];
      LEST[h]<-(LEST[h]+1);
"Step-down fixed"
     p<-1; # to break the while loop
     j<-5;
     while ((j>=1)&(p>0)) {
```

```
   M<-(y[1]-y[6]-lee[5]*(2^0.5));
 for (k in 1:j) {
      for (h in 1:k) {
         w<-(k-h+1);
         F<-(A[h,k]-(lee[5])*((2/w)^0.5));
         if (F>M) {M<-F}
         }
   }
  if (M>=0) {j<-(j-1)} else {p<-0}
}
LDFMED[l]<-(j+1);
h<-LDFMED[l];
LDFEST[h]<-(LDFEST[h]+1);
"WILM PROCEDURE"
r<-1;
for (contrj in 0:4) {
   j<-(5-contrj);
   FMax<-1;
   for (u in 1:j) {
      FMin<-1;
      for (v in j:5) {
         summ<-0;
         for (i in u:v) {
            summ<-(summ+y[i]);
         }
         trymin<-(summ/(v-u+1));
         if (FMin==1) {FMin<-0;curmin<-trymin}
         else {
            if (trymin<curmin) {curmin<-trymin;}
         }
      }
      if (FMax==1) {FMax<-0;curmax<-curmin}
```

```
         else {
            if (curmin>curmax) {curmax<-curmin;}
         }
      }
      if (((curmax-y[6])/sqrt(2))<wilm[j]) {
         r<-j+1;
         break;
      }
   }
   WMED[l]<-r;
   WEST[r]<-(WEST[r]+1);
   "KORN PROCEDURE"
   j<-5;
   p<-1;
   while ((j>=1)&(p>0)) {
      M<-(y[1]-y[6]-(korn[j]*(2^0.5))); f<-1;
      for (i in 1:j) {
         z[i]<-(y[i]-y[6]-(korn[j]*(2^0.5)));
         if (z[i]>M) {M<-z[i]; f<-i}
         }
      if (M>=0) {j<-(f-1)} else {p<-0}
   }
   KMED[l]<-(j+1);

   h<-KMED[l];
   KEST[h]<-(KEST[h]+1);
}
for ( i in 1:6) {
   LEXP<-(LEXP+i*LEST[i]);
   LDFEXP<-(LDFEXP+i*LDFEST[i]);
}
LEXP<-(LEXP/max1);
```

```
LDFEXP<-(LDFEXP/max1);
LBIAS<-(LEXP-T);
LDFBIAS<-(LDFEXP-T);
if (T>1) {
   for (f in 1:(T-1)){
      E<-(E+LEST[f]);
      E1<-(E1+LDFEST[f]);
   }
}
for(f in (T+1):6) {
               E2<-(E2+LEST[f]);
               E4<-(E4+LDFEST[f]);
               }
LFWE<-(E/max1);
LDFFWE<-(E1/max1);
LLACK<-(E2/max1);
LDFLACK<-(E4/max1);
LPOWER<-(LEST[T]/max1);
LDFPOWER<-(LDFEST[T]/max1);
print("Delta"); print(Delta);
print("True MED"); print(T);
print("LEE");
print("LEST");
print(LEST);
print("LEXP");
print(LEXP);
print("LBIAS");
print(LBIAS);
print("LFWE");
print(LFWE);
print("LPOWER");
print(LPOWER);
```

```
   print("LLACK");
   print(LLACK);
print("LEE DOWN FIXED");
   print(LDFEST);
   print(LDFEXP);
   print(LDFBIAS);
   print(LDFFWE);
   print(LDFPOWER);
   print(LDFLACK);
   for ( i in 1:6) {
      KEXP<-(KEXP+i*KEST[i]);
   }
   KEXP<-(KEXP/max1);
   KBIAS<-(KEXP-T);
   if (T>1) {
      for (f in 1:(T-1)){
         G<-(G+KEST[f])
      }
   }
   for (f in (T+1):6) {
                E5<-(E5+KEST[f])
                }
   KFWE<-(G/max1);
   KLACK<-(E5/max1);
   KPOWER<-(KEST[T]/max1);
   print("KORN");
   print(KEST);
   print(KEXP);
   print(KBIAS);
   print(KFWE);
   print(KPOWER);
   print(KLACK);
```

```
   for ( i in 1:6) {
      WEXP<-(WEXP+i*WEST[i]);
   }
   WEXP<-(WEXP/max1);
   WBIAS<-(WEXP-T);
   Q<-0;
   if (T>1) {
      for (f in 1:(T-1)) {
         Q<-(Q+WEST[f])
      }
   }
   for (f in (T+1):6) {
                  E6<-(E6+WEST[f])
                  }
   WFWE<-(Q/max1);
   WLACK<-(E6/max1);
   WPOWER<-(WEST[T]/max1);
   print("WILM");
   print(WEST);
   print(WEXP);
   print(WBIAS);
   print(WFWE);
   print(WPOWER);
   print(WLACK);
}
```

APPENDIX C

R CODE FOR RECURSIVE ALGORITHM FOR IDENTIFYING THE POPULATION MINIMUM EFFICACIOUS DOSE SET FOR A GIVEN RESPONSE MATRIX

The notation is given in Section 3.3.2, with the only changes: matrix F is substituted by `F1` and matrix F' is substituted by `F2`. The MeD-matrix is denoted by `MED`.

```
# Recursive Function for step 4
recur<-function(W,A,B) {
result<-matrix(0,A,B);
l1<-0;l2<-0; m1<-0;m2<-0;
# looking for (l1,l2)
for(a in 1:A) {
              for(b in 1:B) {
              if(W[a,b]==1) {l1<-a;l2<-b;break}
                            }
              if(l1>0) break
              }
# looking for (m1,m2)
for(b in 1:B) {
              for (a in 1:A){
              if(W[a,b]==1) {m1<-a;m2<-b;break}
                            }
              if(m1>0) break
              }
```

```
                if(m1>0) { # non-zero entries present
                if(((m1-l1)>1)&&((l2-m2)>1)) {
     # another run is needed
                W1<-matrix(,m1-l1-1,l2-m2-1);
                for(a in 1:(m1-l1-1)) {
                for(b in 1:(l2-m2-1)) {
                W1[a,b]<-W[l1+a,m2+b]
                                        }
                                        }
                result1<-recur(W1,m1-l1-1,l2-m2-1);
                for(a in 1:(m1-l1-1)) {
                for(b in 1:(l2-m2-1)) {
                result[l1+a,m2+b]<-result1[a,b];
                                                }
                }
                }
     result[l1,l2]<-1;
     result[m1,m2]<-1;
           }
     result
     } #recur
     # The Main Function of the Algorithm of Identifying the
Population MeD-set
     algorithm<-function(M,K,N) {

     #STEP 1. Constructing the Indicator Effect Matrix (IEM)
                IEM<-matrix(0,K,N);
                for(k in 1:K) {
                                for(n in 1:N) {
          if(M[k+1,n+1]>max(M[1,n+1],M[k+1,1])) IEM[k,n]<-1
else IEM[k,n]<-0;
```

```
                                                    }
                                              }

     #STEP 2. Constructing the Matrix F1 (MED for fixed doses
of drug A)
     F1<-matrix(0,K,N);
     for(k in 1:K) {
                    n<-1;
                    if(IEM[k,n]==1) F1[k,n]<-1;
                    while((n<N)&&(IEM[k,n]==0)) {
                    n<-n+1;
                    if(IEM[k,n]==1) F1[k,n]<-1;
                                                  }
                   }
     #STEP 3. Constructing the matrix F2 (MED for fixed doses
of drug B)
     F2<-matrix(0,K,N);
     for(n in 1:N) {
                    k<-1;
                    if(F1[k,n]==1) F2[k,n]<-1;
                    while((k<K)&&(F1[k,n]==0)) {
                    k<-k+1;
                    if(F1[k,n]==1) F2[k,n]<-1;
                                                 }
                   }

     #STEP 4.
     MED<-recur(F2,K,N);
     MED
     }  #algorithm
```

To illustrate how the above algorithm works we have considered the following test procedure. It generates a 6×6 matrix `M` of standard normal responses and applies the algorithm.

```
#TEST
K<-5
N<-5
M<-matrix(0,K+1,N+1)
for (k in 1:(K+1)) {
  for (n in 1:(N+1)) {
     M[k,n]<-rnorm(1, mean=0, sd=1)
                    }
                  }
M
Algorithm(M,K,N)
```

The results of this test are given below. The generated matrix of population responses is given by

-0.2799251	-1.6213077	0.1028200	1.7159707	2.4321761	-3.1145317
0.5636922	0.0620111	-1.9548860	-0.5171184	0.1749682	0.1311853
-0.3697043	0.6650670	0.9565387	0.2911767	0.7744501	0.0178395
0.4097249	-0.8624360	0.3079037	0.2320736	-0.1868923	-1.4978265
-0.3652390	-1.5065386	1.4905473	-1.9176413	-0.3723691	-0.3778648
0.5745144	0.5821338	1.3091003	1.4995865	0.5130860	0.4909296

Then, the indicator gain matrix obtained at the 1st step, together with the matrices F (`F1`) and F' (`F2`) are given by the following matrices, respectively:

0 0 0 0 0
1 1 0 0 1
0 0 0 0 0
0 1 0 0 0
1 1 0 0 0

0 0 0 0 0
1 0 0 0 0
0 0 0 0 0
0 1 0 0 0
1 0 0 0 0

0 0 0 0 0
1 0 0 0 0
0 0 0 0 0
0 1 0 0 0
0 0 0 0 0

And the final MeD-matrix is obtained as

0 0 0 0 0
1 0 0 0 0
0 0 0 0 0
0 0 0 0 0
0 0 0 0 0

APPENDIX D

ON OBTAINING THE ADDITIONAL CRITICAL VALUES FOR THE AVE TEST STATISTIC

In order to prove Theorem 4.2 of Section 4.2, we need to generalize the method, introduced by Hung *et al.* (1993) for obtaining the critical values for the AVE test. Let I be the specified non-rectangular design of the form (4.4). Assume that $n'_{(1)} \geq 2$ and n' is the total number of drug combinations under consideration.

Since for non-rectangular design (4.4)

$$T_A = \frac{\sum_{(i.j)\in I} \hat{\theta}_{ij}}{n' \cdot \hat{\sigma}},$$

the power function of the AVE test statistic is given by

$$P\left(T_A > c \middle| \widehat{\delta}, \widehat{\theta}\right) = \int_0^{\infty} \int_{R^P} \Phi(h_0(u,v,w;\widehat{\delta},\widehat{\theta})) \cdot \prod_{i=1}^{k'} \varphi(u_i) \prod_{j=1}^{n'_{(1))}} \varphi(v_j) du_1 ... du_{k'} dv_1 ... dv_{n'_{(1)}} dQ(w),$$

where $P = k' + n'_{(1)}$,

$$h_0(u,v,w;\widehat{\delta},\widehat{\theta}) = -\sqrt{n \cdot n'}\left(cw - \frac{\theta_{AVE}}{\sigma}\right) - \frac{\sum_{(i,j)\in I} \left[u_i + v_j - \sqrt{n}\left|\delta_{ij}\right| + \left|u_i - v_j + \sqrt{n}\delta_{ij}\right|\right]}{2\sqrt{n'}} \text{ and}$$

$$\theta_{AVE} = \frac{\sum_{(i,j)\in I} \theta_{ij}}{n'}.$$

The power function depends on the parameters θ_{ij} only through θ_{AVE} and for any given $\hat{\delta}$ with $\delta_{ij} \in (-\infty, \infty)$ and any fixed number $c \in (-\infty, \infty)$, it is increasing in θ_{AVE}. So the significance level is defined as

$$Significance\ \ Level \equiv \sup_{all\ \delta_{ij} \in (-\infty,\infty)} P\left(T_A > c \middle| \hat{\delta}, \hat{\theta} = 0\right).$$

Similar to Appendix B of Hung *et al.* (1993), it can be shown that for any given θ_{AVE} and c, the power function is increasing in each $\delta_{ij} \in (0, \infty)$ and is decreasing in each $\delta_{ij} \in (-\infty, 0)$ when the remaining δ_{lm} are fixed. To prove it, consider the following cases.

Case D.1. $\delta_{11} \in (0, \infty)$ and $u_i - v_j + \sqrt{n}\delta_{11} > 0$, then

$$\begin{aligned} & h_0(u, v, w; \hat{\delta}, \hat{\theta} = 0) \\ &= -\sqrt{n \cdot n'}(cw) - \frac{\sum_{(i,j) \in I} \left[u_i + v_j - \sqrt{n}\left|\delta_{ij}\right| + \left|u_i - v_j + \sqrt{n}\delta_{ij}\right| \right]}{2\sqrt{n'}} \\ &\quad - \frac{u_1 + v_1 - \sqrt{n}\delta_{11} + u_1 - v_1 + \sqrt{n}\delta_{11}}{2\sqrt{n'}} \\ &= -\sqrt{n \cdot n'}(cw) - \frac{\sum_{\substack{i=1,\ldots,K;\ j=1,\ldots,N \\ (i,j) \neq (1,1),(K,N)}} \left[u_i + v_j - \sqrt{n}\left|\delta_{ij}\right| + \left|u_i - v_j + \sqrt{n}\delta_{ij}\right| \right]}{2\sqrt{n'}} - \frac{u_1}{\sqrt{n'}} \\ &= h_0(u, v, w, \hat{\delta}_{(11)}, \hat{\theta} = 0) - \frac{u_1}{\sqrt{n'}}, \text{ where } \hat{\delta}_{(11)} \text{ denotes the vector of } \delta_{ij} \text{ excluding } \delta_{11}. \end{aligned}$$

Case D.2. $\delta_{11} \in (0, \infty)$ and $u_i - v_j + \sqrt{n}\delta_{11} \leq 0$, then

$$\begin{aligned} & h_0(u, v, w; \hat{\delta}, \hat{\theta} = 0) \\ &= -\sqrt{n \cdot n'}(cw) - \frac{\sum_{(i,j) \in I} \left[u_i + v_j - \sqrt{n}\left|\delta_{ij}\right| + \left|u_i - v_j + \sqrt{n}\delta_{ij}\right| \right]}{2\sqrt{n'}} \\ &\quad - \frac{u_1 + v_1 - \sqrt{n}\delta_{11} - u_1 + v_1 - \sqrt{n}\delta_{11}}{2\sqrt{n'}} \\ &= h_0(u, v, w, \hat{\delta}_{(11)}, \hat{\theta} = 0) - \frac{v_1 - \sqrt{n}\delta_{11}}{\sqrt{n'}}. \end{aligned}$$

Combining the final results of Cases D.1 and D.2 for $\delta_{11} \in (0, \infty)$, we obtain

$$P\left(T_A > c\middle|\hat{\delta},\hat{\theta}=0\right)=$$

$$\int_0^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{u_1+\sqrt{n}\delta_{11}}\int_{R^T}\Phi\left[h_0(u,v,w,\hat{\delta}_{(11)},0)-\frac{u_1}{\sqrt{n'}}\right]\prod_{i=2}^{k'}\varphi(u_i)\prod_{j=2}^{n'_{(1)}}\varphi(v_j)d\hat{u}_{(1)}d\hat{v}_{(1)}dv_1du_1dQ(w)$$

$$+\int_0^{\infty}\int_{-\infty}^{\infty}\int_{u_1+\sqrt{n}\delta_{11}}^{\infty}\int_{R^T}\Phi\left[h_0(u,v,w,\hat{\delta}_{(11)},0)-\frac{v_1-\sqrt{n}\delta_{11}}{\sqrt{n'}}\right]\prod_{i=2}^{k'}\varphi(u_i)\prod_{j=2}^{n'_{(1)}}\varphi(v_j)d\hat{u}_{(1)}d\hat{v}_{(1)}dv_1du_1dQ(w),$$

where $T=k'+n'_{(1)}-2$ $d\hat{u}_{(1)}=du_2....du_{k'}$, $d\hat{v}_{(1)}=dv_2....dv_{n'_{(1)}}$.

Then, the derivative with respect to $\delta_{11}\in(0,\infty)$ is given by

$$\frac{dP(T_A>c\left|\hat{\delta},\hat{\theta}=0\right.)}{d\delta_{11}}=\int_0^{\infty}\int_{-\infty}^{\infty}\int_{u_1+\sqrt{n}\delta_{11}}^{\infty}\int_{R^T}\varphi[t]\cdot\sqrt{\frac{n}{n'}}\prod_{i=2}^{k'}\varphi(u_i)\prod_{j=2}^{n'_{(1)}}\varphi(v_j)d\hat{u}_{(1)}d\hat{v}_{(1)}dv_1du_1dQ(w)>0,$$

where $t=h_0(u,v,w,\hat{\delta}_{(11)},0)-\frac{v_1-\sqrt{n}\delta_{11}}{\sqrt{n'}}$, so that $P\left(T_A>c\middle|\hat{\delta},\hat{\theta}=0\right)$ is increasing in $\delta_{11}\in(0,\infty)$ when the remaining δ_{lm} are fixed.

Since $\delta_{11}\in(0,\infty)$ is arbitrary, the same conclusion holds for any other $\delta_{ij}\in(0,\infty)$. Similarly, it can be shown that the derivative is negative in the case of each $\delta_{ij}\in(-\infty,0)$, when the remaining δ_{lm} are fixed. So we can conclude that the power function is increasing in each $\delta_{ij}\in(0,\infty)$ and is decreasing in each $\delta_{ij}\in(-\infty,0)$ when the remaining δ_{lm} are fixed.

Therefore, since any bounded monotone function converges, the significance level can be evaluated as

$$Significance\ \ Level\equiv\max\left\{\lim_{|\delta_{ij}|\to\infty}P(T_A>c\left|\hat{\delta},\hat{\theta}=0\right.),\ \ i=1,...,k';\ \ j=1,...,n'_{(1)}\right\}.$$

To evaluate the limit, let us first consider the limit when a single $|\delta_{11}|\to\infty$ by considering separately the cases when $\delta_{11}\to\infty$ and $\delta_{11}\to-\infty$.

<u>Case D.3</u>. $\delta_{11}\to\infty$

Since

$$\int_0^{\infty}\int_{-\infty}^{\infty}\int_{u_1+\sqrt{n}\delta_{11}}^{\infty}\int_{R^T}\Phi\left[h_0(u,v,w,\hat{\delta}_{(11)},0)-\frac{v_1-\sqrt{n}\delta_{11}}{\sqrt{n'}}\right]\prod_{i=2}^{k'}\varphi(u_i)\prod_{j=2}^{n'_{(1)}}\varphi(v_j)d\hat{u}_{(1)}d\hat{v}_{(1)}dv_1du_1dQ(w)\to 0,$$

$$\lim_{\delta_{11}\to\infty}P\left(T_A>c\middle|\hat{\delta},\hat{\theta}=0\right)$$

$$= \int_0^{\infty} \int_{-\infty}^{\infty} \int_{R^{T2}}^{1} \int^{\infty} \Phi\left[h_0(u,v,w,\hat{\delta}_{(11)},0) - \frac{u_1}{\sqrt{n'}} \right] \prod_{i=2}^{k'} \varphi(u_i) \prod_{j=2}^{n'_{(1)}} \varphi(v_j) d\hat{u}_{(1)} d\hat{v}_{(1)} dv_1 du_1 dQ(w)$$

$$= \int_0^{\infty} \int_{R^P} \Phi\left[h_0(u,v.w.\hat{\delta}_{(11)},0) - \frac{u_1}{\sqrt{n'}} \right] \prod_{i=1}^{k'} \varphi(u_i) \prod_{j=1}^{n'_{(1)}} \varphi(v_j) du_1 ... du_{k'} dv_1 ... dv_{n'_{(1)}} dQ(w) .$$

<u>Case D.4</u>. $\delta_{11} \to -\infty$

Then, the first term of $P\left(T_A > c \middle| \hat{\delta}, \hat{\theta} = 0\right)$ goes to zero and the second term becomes

$$\int_0^{\infty} \int_{R^{k'+n'(1)}} \Phi(h_0(u,v,w,\hat{\delta}_{(11)},0) - \frac{v_1}{\sqrt{n'}}) \cdot \prod_{i=1}^{k'} \varphi(u_i) \prod_{j=1}^{n'_{(1)}} \varphi(v_j) du_1 ... du_{k'} dv_1 ... dv_{n'_{(1)}} dQ(w)$$

Combining the above cases we can conclude that

$$\lim_{|\delta_{11}| \to \infty} P\left(T_A > c \middle| \hat{\delta}, \hat{\theta} = 0\right)$$

$$= \int_0^{\infty} \int_{R^P} \Phi\left[h_0(u,v.w.\hat{\delta}_{(11)},0) - \frac{\pi_{11} u_1 + (1-\pi_{11}) v_1}{\sqrt{n'}} \right] \prod_{i=1}^{k'} \varphi(u_i) \prod_{j=1}^{n'_{(1)}} \varphi(v_j) du_1 ... du_{k'} dv_1 ... dv_{n'_{(1)}} dQ(w) ,$$

where

$$\pi_{11} = \begin{cases} 1, & if \quad \delta_{11} \to \infty \\ 0, & if \quad \delta_{11} \to -\infty . \end{cases}$$

Then, the final limit when all $|\delta_{ij}| \to \infty$ can be obtained as

$$\lim_{|\delta_{ij}| \to \infty} P(T_A > c \middle| \hat{\delta}, \hat{\theta} = 0) = \lim_{\substack{|\delta_{ij}| \to \infty \\ (i,j) \neq (1,1)}} \left(\lim_{|\delta_{ii}| \to \infty} P(T_A > c \middle| \hat{\delta}, \hat{\theta} = 0) \right)$$

$$= \int_0^{\infty} \int_{R^P} \Phi\left[\lim_{\substack{|\delta_{ij}| \to \infty \\ (i,j) \neq (1,1)}} h_0(u,v.w.\hat{\delta}_{(11)},0) - \frac{\pi_{11} u_1 + (1-\pi_{11}) v_1}{\sqrt{n'}} \right] \prod_{i=1}^{k'} \varphi(u_i) \prod_{j=1}^{n'_{(1)}} \varphi(v_j) du_1 ... du_{k'} dv_1 ... dv_{n'_{(1)}} dQ(w)$$

where

$$\lim_{\substack{|\delta_{ij}| \to \infty \\ (i,j) \neq (1,1)}} h_0(u,v,w,\hat{\delta}_{(11)},0)$$

$$= \lim_{\substack{|\delta_{ij}| \to \infty \\ (i,j) \neq (1,1)}} \left(-\sqrt{n \cdot n'} cw \right) - \lim_{\substack{|\delta_{ij}| \to \infty \\ (i,j) \neq (1,1)}} \frac{\sum_{\substack{(i,j) \in I \\ (i.j) \neq (1,1)}} \left[u_i + v_j - \sqrt{n} \middle| \delta_{ij} \middle| + \middle| u_i - v_j + \sqrt{n} \delta_{ij} \middle| \right]}{2\sqrt{n'}}$$

$$= -\sqrt{n \cdot n'}cw - \begin{cases} \dfrac{\sum\limits_{\substack{(i,j)\in I \\ (i,j)\neq(1,1)}} (u_i + v_j + u_i - v_j)}{2\sqrt{n'}}, & if \ \ all \ \ \delta_{ij} \to \infty, \\ \dfrac{\sum\limits_{\substack{(i,j)\in I \\ (i,j)\neq(1,1)}} u_i + v_j + v_j - u_i}{2\sqrt{n'}}, & if \ \ all \ \ \delta_{ij} \to -\infty \end{cases}$$

That is,

$$\lim_{\substack{|\delta_{ij}|\to\infty \\ (i,j)\neq(1,1)}} h_0(u,v,w,\hat{\delta}_{(11)},0) = \left(-\sqrt{n \cdot n'}cw\right) - \frac{\sum\limits_{\substack{(i,j)\in I \\ (i.j)\neq(1,1)}} \left[\pi_{ij}u_i + (1-\pi_{ij})v_j\right]}{\sqrt{n'}},$$

where

$$\pi_{ij} = \begin{cases} 1, & if \ \ \delta_{ij} \to \infty \\ 0, & if \ \ \delta_{ij} \to -\infty. \end{cases}$$

And the limit of $P(T_A > c|\hat{\delta}, \hat{\theta} = 0)$ as all $|\delta_{ij}| \to \infty$ becomes

$$\lim_{|\delta_{ij}|\to\infty} P(T_A > c|\hat{\delta}, \hat{\theta} = 0) = \int_0^\infty \int_{R^P} \Phi[q] \prod_{i=1}^{k'} \varphi(u_i) \prod_{j=1}^{n'_{(1)}} \varphi(v_j) du_1 ... du_{k'} dv_1 ... dv_{n'_{(1)}} dQ(w) \text{ , where}$$

$$q = \left(-\sqrt{n \cdot n'}cw\right) - \frac{\sum\limits_{(i,j)\in I} \left[\pi_{ij}u_i + (1-\pi_{ij})v_j\right]}{\sqrt{n'}}.$$

To simplify the limit expression, let us consider the numerator of the second term in q, that is

$$\sum_{(i,j)\in I} \left[\pi_{ij}u_j + (1-\pi_{ij})v_j\right] = \sum_{j=1}^{n'_{(1)}} \pi_{1j}u_1 + ... + \sum_{j=1}^{n'_{(k'-1)}} \pi_{k'-1\ j}u_{k'-1} + \sum_{j=1}^{n'_{(k)}} \pi_{k'j}u_{k'}$$

$$+ \sum_{i:(i,1)\in I} (1-\pi_{i1})v_1 + ... + \sum_{i:(i,n'_{(2)})\in I} (1-\pi_{in'_{(2)}})v_{n'_{(2)}} + \sum_{i:(i,n'_{(1)})\in I} (1-\pi_{in'_{(1)}})v_{n'_{(1)}}$$

$$= m_1 z_1 + ... + m_{k'} z_{k'} + m_{k'+1} z_{k'+1} + ... + m_{k'+n'_{(1)}} z_{k'+n'_{(1)}} \text{ , where}$$

$$m_s = \begin{cases} \sum\limits_{j=1}^{n'_{(s)}} \pi_{sj}, & s = 1,...k' \\ \sum\limits_{i:(i,s)\in I} (1-\pi_{i,s-k'}), & s = k'+1,..., P = k'+n'_{(1)} \end{cases}$$

and

$$z_s = \begin{cases} u_s, & s = 1,...k' \\ v_{s-k'}, & s = k'+1,..., P = k'+n'_{(1)}. \end{cases}$$

So that, q can be written as

$$q = \left(-\sqrt{n \cdot n'}cw\right) - \frac{\sum_{s=1}^{P} m_s z_s}{\sqrt{n'}}.$$

Since z_s, $s = 1,..., P$ are standard normal random variables,

$$\sum_{s=1}^{P} m_s z_s \sim N\left(0,\ Var = \sum_{s=1}^{P} m_s^2\right) \text{ and } x = \frac{\sum_{s=1}^{P} m_s z_s}{\sqrt{\sum_{s=1}^{P} m_s^2}} \sim N(0,\ 1),$$

the limit can be expressed as

$$\lim_{|\delta_{ij}|\to\infty} P(T_A > c \mid \hat{\delta}, \hat{\theta} = 0)$$

$$= \int_0^{\infty} \int_{R^P} \Phi\left[-\sqrt{n \cdot n'}cw - \frac{1}{\sqrt{n'}} \sum_{s=1}^{P} m_s z_s\right] \prod_{s=1}^{P} \varphi(z_s) dz_1 ... dz_P dQ(w)$$

$$= \int_0^{\infty} E\left(\Phi\left[-\sqrt{n \cdot n'}cw - \frac{1}{\sqrt{n'}} \sum_{s=1}^{P} m_s z_s\right]\right) dQ(w)$$

$$= \int_0^{\infty} E\left(\Phi\left[-\sqrt{n \cdot n'}cw - \frac{1}{\sqrt{n'}} \cdot x\sqrt{\sum_{s=1}^{P} m_s^2}\right]\right) dQ(w)$$

$$= \int_0^{\infty} \int_{-\infty}^{\infty} \Phi\left[-\sqrt{n \cdot n'}cw - \sqrt{\frac{\sum_{s=1}^{P} m_s^2}{KN-1}} \cdot x\right] \varphi(x) dx dQ(w).$$

Let $G(t) = \int_0^{\infty} \int_{-\infty}^{\infty} \Phi\left[-\sqrt{n \cdot n'}cw - tx\right] \varphi(x) dx dQ(w)$, where

$$t = \sqrt{\frac{\sum_{s=1}^{P} m_s^2}{n'}} > 0, \text{ then}$$

$G(t)$ can be written as

$$G(t) = \int_0^{\infty}\int_{-\infty}^{0} \Phi\left[-\sqrt{n\cdot n'}cw - tx\right]\varphi(x)dxdQ(w) + \int_0^{\infty}\int_0^{\infty} \Phi\left[-\sqrt{n\cdot n'}cw - tx\right]\varphi(x)dxdQ(w)$$

Next, we substitute a new variable $x' = -x$ in the first double integral, then the limits of integration change from $(-\infty,0)$ to $(\infty,0)$, since the density function of the standard normal random variable is even, $\varphi(x) = \varphi(x')$ and $dx = -dx'$, so by switching the limits of integration to $(0,\infty)$ we obtain

$$G(t) = \int_0^{\infty}\int_0^{\infty} \Phi\left[-\sqrt{n\cdot n'}cw + tx'\right]\varphi(x')dx'dQ(w) + \int_0^{\infty}\int_0^{\infty} \Phi\left[-\sqrt{n\cdot n'}cw - tx\right]\varphi(x)dxdQ(w)$$

$$= \int_0^{\infty}\int_0^{\infty} \left(\Phi\left[-\sqrt{n\cdot n'}cw - tx\right] + \Phi\left[-\sqrt{n\cdot n'}cw + tx\right]\right)\varphi(x)dxdQ(w)$$

Next, our goal is to show that $G(t)$ is an increasing function in t, then the significance level can be obtained as the value of $G(t)$, evaluated at the largest possible t. To show that, we will consider the derivative of $G(t)$ with respect to t for any fixed $c > 0$ and any given $w > 0$, and prove that it is positive.

Since the outer integral does not depend on t, the order of integration and differentiation can be changed and we obtain:

$$G'_t(t) = \int_0^{\infty}\int_0^{\infty} \left((-x)\varphi\left[-\sqrt{n\cdot n'}cw - tx\right] + x\varphi\left[-\sqrt{n\cdot n'}cw + tx\right]\right)\varphi(x)dxdQ(w)$$

$$= \int_0^{\infty}\int_0^{\infty} \left(\varphi\left[\sqrt{n\cdot n'}cw - tx\right] - \varphi\left[\sqrt{n\cdot n'}cw + tx\right]\right)x\varphi(x)dxdQ(w) > 0,$$

because for $x > 0,\ c > 0,\ w > 0,\ t > 0$, $\sqrt{n\cdot n'}cw + tx > 0$. If $\sqrt{n\cdot n'}cw - tx > 0$ then $\sqrt{n\cdot n'}cw - tx < \sqrt{n\cdot n'}cw + tx$ and $\varphi(\sqrt{n\cdot n'}cw - tx) > \varphi(\sqrt{n\cdot n'}cw + tx)$. If $\sqrt{n\cdot n'}cw - tx < 0$ then $\left|\sqrt{n\cdot n'}cw - tx\right| < \sqrt{n\cdot n'}cw + tx$ and $\varphi(\sqrt{n\cdot n'}cw - tx) > \varphi(\sqrt{n\cdot n'}cw + tx)$. So the function difference in the double integral is always positive, which proves that $G_t'(t) > 0$ and, hence, $G(t)$ is increasing in t.

The last fact allows us to claim that the significance level can be obtained as the value of $G(t)$, evaluated at the largest possible t, subject to the constrains, specified above. This completes the proof of the theorem.

APPENDIX E

THE OPTIMIZATION PROBLEM OF SECTION 4.2

In Section 4.2 we formulate Theorem 4.2 which provides the equation for obtaining the critical values for the non-rectangular designs of the form (4.4). The proof of Theorem 4.2 is given in Appendix D. We also have mentioned that before the result can be applied, one should solve the following optimization problem.

Problem. Let $k' \le K$, $n'_{(k')} \le n'_{(k'-1)} \le \ldots \le n'_{(1)} \le N$ with $n'_{(k')} < n'_{(1)}$. Consider the non-rectangular design of the form $I \equiv \{(i, j), j = 1,\ldots, n'_{(i)}, i = 1,\ldots, k'\}$, where $n' = n'(k') = \sum_{i=1}^{k'} n'_{(i)}$ denotes the cardinality of I and $P = k' + n'_{(1)}$. Then, the problem is to find the maximum value

$$t_{\max} = \max \sqrt{\frac{\sum_{s=1}^{P} m_s^2}{n'}},$$

where the non-negative integers m_s, $s = 1,\ldots, P$ are subject to the following constraints:

$$m_s \le n'_{(s)}, \quad s = 1,\ldots, k';$$
$$m_s \le \left|\{(i, s-k') : (i, s-k') \in I\}\right|, \quad s = k'+1,\ldots, P$$
$$\sum_{s=1}^{P} m_s = n',$$

where $|A|$ denotes the cardinality of a set A.

We provide the solution to this problem for the specific form of the non-rectangular design, when a single highest dose combination is excluded from the rectangular design. This

result is presented in the Example 4.1 of Section 4.2 and states that for a non-rectangular design of the form, $I_0 = G \setminus \{(K,N)\} \equiv \{\{(i,j), i = 1,...,K,\ j = 1,...,N\} \setminus \{(K,N)\}\}$, the maximum value

$$t_{\max} = \max \sqrt{\frac{\sum_{s=1}^{P} m_s^2}{n'}},$$

where m_s, $s = 1,...,P$ are non-negative integers subject to constraints:

$$m_s \le n'_{(s)},\ \ s = 1,...,k';$$
$$m_s \le \left|\{(i, s-k') : (i, s-k') \in I\}\right|,\ \ s = k'+1,...,P$$
$$\sum_{s=1}^{P} m_s = n',$$

equals $t_{\max} = \sqrt{\frac{ab^2 - 2b + 1}{n'}}$, where $a = \min\{K,\ N\}$, $b = \max\{K,\ N\}$ and $n' = KN - 1$.

We now prove this fact. Since the non-rectangular design $I_0 \equiv I$ with $n' = KN - 1$, $P = K + N$, $k' = K$, $n'_{(s)} = N$ for $s \le K - 1$ and $n'_{(K)} = N - 1$, the above constraints reduce to the following conditions:

$$m_s \le N,\ \ s = 1,...,K-1;$$
$$m_K \le N - 1$$
$$m_s \le K,\ \ s = K+1,...,K+N-1$$
$$m_{K+N} \le K - 1$$
$$\text{and } \sum_{s=1}^{K+N} m_s = KN - 1.$$

Next, consider the case when $\max\{K,\ N\} = K$ and $\min\{K,\ N\} = N$, so $N \le K$. Note that t is increasing in each m_s, and there are $N - 1$ numbers

$$m_{K+1} \le K,\ m_{K+2} \le K,...,\ m_{K+N-1} \le K,$$

that can be evaluated at the maximum possible value, K. The sum of these $N - 1$ numbers is equal to $K(N-1) = KN - K < KN - 1$ for $K > 1$. Hence, to satisfy the condition

$$\sum_{s=1}^{K+N} m_s = KN - 1,$$

we should consider additional m_s's, as few of them as possible, which sum to S, such that $K(N-1)+S=KN-1$, *i.e.*, $S=K-1$. There is a single number $m_{K+N} \le K-1$ which can be evaluated at its maximum possible value and the rest of the numbers are taken to be zero. Then

$$\sum_{s=1}^{K+N} m_s = \sum_{s=1}^{K} m_s + \sum_{s=K+1}^{K+N-1} m_s + m_{K+N} = 0 + K(N-1) + K - 1 = KN - 1.$$

Numbers, chosen in such a way provide the combination which maximizes t and

$$t_{\max} = \sqrt{\frac{K^2(N-1)+(K-1)^2}{KN-1}} = \sqrt{\frac{K^2N-2K+1}{KN-1}}.$$

Similarly, in the case when $\max\{K,\ N\} = N$ and $\min\{K,\ N\} = K$ we obtain

$$t_{\max} = \sqrt{\frac{N^2K-2N+1}{KN-1}}.$$

The results of both cases can be combined by letting $a = \min\{K,\ N\}$ and $b = \max\{K,\ N\}$ to obtain the final expression in the form

$$t_{\max} = \sqrt{\frac{ab^2-2b+1}{n'}}, \text{ where } n' = KN-1.$$

■

In Section 4.4 we provide the equation for obtaining the critical values for the AVE test for testing $H_0^{(4')}: \theta_{11} = \theta_{12} = \theta_{13} = \theta_{21} = 0$, which corresponds to the design

$$I' = \{(1,1),\ (1,2),\ (1,3),\ (2,1)\}.$$

Let us illustrate how the result is obtained. The design $I' \equiv I$ with $n' = 4$, $k' = 2,\ n'_{(1)} = 3,\ n'_{(2)} = 1$, then $P = k' + n'_{(1)} = 5$. So the problem becomes: to maximize

$$t = \sqrt{\frac{\sum_{s=1}^{5} m_s^2}{4}},$$

where $m_1 \le 3,\ m_2 \le 1,\ m_3 \le \left|\{(i,1) : (i,1) \in I'\}\right| = \left|\{(1,1),\ (2,1)\}\right| = 2$, $m_4 \le \left|\{(i,2) : (i,2) \in I'\}\right| = \left|\{(1,2)\}\right| = 1$ and $m_5 \le \left|\{(i,3) : (i,3) \in I'\}\right| = \left|\{(1,3)\}\right| = 1$, such that

$$\sum_{s=1}^{5} m_s = 4.$$

Since t is increasing in each m_s, it is maximized when a single m_s (or a few of them) is evaluated at its maximum and the rest are taken to be zero. First, we rewrite the non-negative integers, m_s's, according to the maximum possible value: $m_1 \le \max_1 = 3,\ m_3 \le \max_3 = 2$, $m_t \le \max_t = 1,\ t = 2,\ 4,\ 5$. Next, since $\max_1 + \max_3 = 5 > 4$ we should decrease the value of m_3 and consider $m_3 = 1$. Then $\max_1 + m_3 = 4$ and we can take $m_1 = 3,\ m_3 = 1$ and $m_t = 0,\ t = 2,\ 4,\ 5$. Thus, we obtain the maximized value

$$t_{\max} = \sqrt{\frac{10}{4}}\,.$$

■

The same method can be used to obtain the solution to the optimization problem for any other non-rectangular design I.

BIBLIOGRAPHY

1. Bartholomew, D.J. (1959a). A Test of Homogeneity for Ordered Alternatives. *Biometrika* 46, 36-48.
2. Bartholomew, D.J. (1959b). Ordered Tests in the Analysis of Variance. *Biometrika* 48, 328-335.
3. CPMP/EWP (2002). Committee for Proprietary Medicinal Products, Efficacy Working Party. Points to Consider on Multiplicity Issues in Clinical Trials, *CPMP/EWP/908/99.* Available at http://www.emea.eu.int/pdfs/human/ewp/090899en.pdf
4. Hellmich, M. and Lehmacher, W. (2005). Closure Procedures for Monotone Bi-factorial Dose-Response Designs. *Biometrics 61, 269-276*.
5. Hung, H.M. (1992). On Identifying a Positive Dose-Response Surface for Combination Agents. *Statistics in Medicine 11, 703-711.*
6. Hung, H.M., Chi, J.Y., and Lipicky, R.J. (1993). Testing for the Existence of a Desirable Dose Combination. *Biometrics 49, 85-94*.
7. Jia, G. (2004). Use of Simultaneous Inference Under Order Restriction, Stepdown Testing Procedure and Stage-Wise Sequential Optimal Design in Clinical Dose Study. *PhD Dissertation, University of Pittsburgh, Department of Statistics.*
8. Korn, E.L. (1982). Confidence Bands for Isotonic Dose-Response Curves. *Applied Statistics 31, 59-63.*
9. Laska, E. M., and Meisner, M. (1989). Testing Whether an Identified Treatment is Best. *Biometrics 45, 1159-1151*.
10. Lee, C.C. (1996). On Estimation for Monotone Dose-Response Curves. *JASA 91, 1110-1119.*
11. Marcus, R., Peritz, E., and Gabriel, K.R. (1976). On Closed Testing Procedures with Special Reference to Ordered Analysis of Variance. *Biometrika 63, 655-660.*
12. Palangio, M., Gilder, W., Keffer, M., Landan, C., Morris, E., Doyle, R., Jiang, J., Damask, M., and de Padova, A. (2000). Dose-Response Effect of Combination Hydrocodone with Ibuprofen in Patients with Moderate to Severe Postoperative Pain. *Clinical Therapeutics, 22, No. 8, 990-1002.*
13. Rom, D.M., Costello, R.J., and Connell, L.T. (1994). On Closed Test Procedures for Dose-Response Analysis. *Statistics in Medicine 18, 1583-1596.*

14. Sampson, A.R., Singh, H., Whitaker, L.R. (2005). Simultaneous Confidence Bands for Isotonic Functions. *Under Review.*

15. Tamhane, A.C., Hochberg, Y., Dunnett, C.W. (1996). Multiple Test Procedures for Dose Finding. *Biometrics 52, 21-37.*

16. U.S. Food and Drug Administration's policy "Fixed-Combination Prescription Drugs for Humans". *Code of Federal Regulations, Title 21, Volume 5* (21 CFR 300.50). Revised April 1, 2005. Available at http://www.washingtonwatchdog.org/documents/cfr/title21/part300.html#300.50

17. Williams, D.A. (1971). A Test for Differences between Treatment Means When Several Dose Levels are Compared with a Zero Dose Control. *Biometrics 27, 103-117.*

18. Williams, D.A. (1972). The Comparison of Several Dose Levels with a Zero Dose Control. *Biometrics 28, 519-531.*

19. Wertheimer, A. I., Morrison, A. (2002). Combination Drugs: Innovation in Pharmacotherapy. *Pharmacology and Therapeutics 27, No. 1, 44-49.*