Published in final edited form as:

Stat Biopharm Res. 2009 ; 1(1): 39–47. doi:10.1198/sbr.2009.0004.

On Identifying Minimum Efficacious Doses in Combination Drug Trials

Julia N. Soulakova and

Department of Statistics, University of Nebraska-Lincoln, 340 Hardin Hall-North, Lincoln, NE 68583-0963 (E-mail: jsoulakova2@unlnotes.unl.edu)

Allan R. Sampson

Department of Statistics, University of Pittsburgh, 2717 Cathedral of Learning, Pittsburgh, PA 15260

Abstract

To determine that a combination drug is effective, the FDA requires demonstration that the combination itself is effective and that each component makes a contribution to the claimed effect. In a study with at least one component known to be effective at the considered doses, these regulatory requirements can be satisfied by showing that the combination is superior to each component. We term such a combination an *efficacious* combination. In a dose-response study involving combination drugs, one of which is known to be effective, we are interested in detecting those combinations which are efficacious, and for which no lower dose combination is also efficacious. We term these combinations the minimum efficacious dose set and our goal is to estimate this set. Our procedure requires first identifying all possible minimum efficacious dose sets and the corresponding hypotheses for a given design. Next, the proper testing order based on a graph representation is established and the hypotheses are tested using the “average” test under the closed testing principle. This procedure is shown to have strong control of overall error rate. The power of this procedure is studied by simulation.

Keywords

Average test; Closed testing; Dose-response; Minimum effective dose; Step-down procedure

1. Introduction

Combination products, also known as fixed-dose combinations, are combinations of two or more active compounds. Often, these active drugs come from two different classes of drugs intended to have the same therapeutic effect. The main rationales for combining drugs are to produce a more efficacious treatment and to simplify the treatment regimens.

In order for a new combination drug to be approved for human use, the U.S. Food and Drug Administration (FDA) requires demonstration 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, one must demonstrate that the combination is effective, that it is superior to placebo, and that the combination is superior to its components.

Our focus in this article is to provide a methodology for analyzing dose-response studies involving combination drugs. For the settings where one component is known to be effective at all doses studied in the combination, we introduce the notion of the set of the minimum efficacious combinations and provide a procedure to estimate this set. In the setting where none

of the components is known to be effective, our procedure can be interpreted as providing the minimum “superior” combinations, where a superior combination is one superior to each component. For ease of presentation, we discuss only the former setting.

Consider a $K \times N$ combination drug trial, where $n(K+1)(N+1)$ subjects are randomly allocated in a balanced fashion, n per group, to one of $(K+1) \times (N+1)$ groups. Subjects of each group receive a combination of drug A at dose level i and drug B at dose level j , where $i = 0$ and $j = 0$ denote the zero doses. We call the preceding design “rectangular.” In Section 3.2 we define another type of design, called “nonrectangular,” which occurs when one or several “highest” doses are omitted from the study. The procedure presented in this article can be applied to either rectangular or to nonrectangular design; however, for simplicity, we state the results for the rectangular case. We note that while typical combination studies might include the combination with both $i = 0$ and $j = 0$, that is, placebo, our procedure does not require the inclusion of this combination. For notation convenience, we suppose that this combination is included.

Let μ_{ij} be the population mean dose–response for combination (i, j) , and define the expected gain from combining the i th dose of drug A with the j th dose of drug B as $\theta_{ij} = \min\{\mu_{ij} - \mu_{i0}, \mu_{ij} - \mu_{0j}\}$, $i = 1, \dots, K$ and $j = 1, \dots, N$. If this gain is positive, Hung, Chi, and Lipicky (1993) termed the combination as superior to each component. It is common to assume that the combination of the two drugs is not expected to be inferior with respect to efficacy to either component alone, that is, all $\theta_{ij} \geq 0$; this is assumed by us throughout. Then any combination with $\theta_{ij} > 0$ is an efficacious combination. Furthermore, let $y_{ij,t}$ be the response of the t th subject in the (i, j) th dose combination group. We assume that $y_{ij,t}$ are independent and $y_{ij,t} \sim N(\mu_{ij}, \sigma^2)$, $i = 0, \dots, K$, $j = 0, \dots, N$ and $t = 1, \dots, n$.

We introduce the notion of a *minimum efficacious dose (MeD)* as an efficacious combination, such that decreasing any component dose leads to a nonefficacious compound, that is, a dose (r, s) is an MeD if all of the following are true:

$$\begin{aligned} \theta_{rs} &> 0; \\ \theta_{ij} &= \theta_{rj} = \theta_{is} = 0, \quad \text{for all } i < r, j < s. \end{aligned}$$

Since the MeD is not necessarily unique, we introduce the notion of the MeD-set, defined as the set of all the MeD's.

2. Background

For the case when only multiple doses of a single compound are being evaluated, Tamhane, Hochberg, and Dunnett (1996) provided an elegant approach to estimating the minimum effective dose (MED). They used a closed family of hypotheses with a completely ordered step-wise testing procedure which strongly controls the family-wise error rate (FWE). The MED is estimated based on the results of the sequence of hypothesis tests. For the combination drug setting, there do not appear to be any analogous approaches to this. There are, however, several articles that deal with various testing aspects for combination drugs; we now summarize these in the following.

Hung et al. (1993) considered a problem of “global superiority,” that is, a problem of identifying whether (or not) there exists at least one combination superior to each component. This problem is stated in terms of the hypotheses $H_0 : \theta_{ij} = 0, \text{ for all } i, j \geq 1$, and $H_a : \theta_{ij} > 0, \text{ for some } i, j \geq 1$, where they assumed all $\theta_{ij} \geq 0$. In order to test these hypotheses, one of the tests they consider is the AVE test, based on the test statistic, T_A , which is given by

$$T_A = \sum_{i=1}^K \sum_{j=1}^N \hat{\theta}_{ij} / (K N \hat{\sigma}), \quad (1)$$

where $\hat{\theta}_{ij}, \hat{\theta}_{ij} = \bar{y}_{ij} - \max(\bar{y}_{i0}, \bar{y}_{0j})$, is an estimator of θ_{ij} . The pooled estimator $\hat{\sigma}^2$ of σ^2 is given by $\hat{\sigma}^2 = \sum_{i=0}^K \sum_{j=0}^N \sum_{t=1}^n (y_{ijt} - \bar{y}_{ij})^2 / v$, where $v = (K+1)(N+1)(n-1)$ so that $v\hat{\sigma}^2/\sigma^2$ has a χ^2 distribution with v degrees of freedom.

Hung et al. (1993) provided the distribution function of AVE test. This distribution function depends on the parameters θ_{ij} 's only through the average value of θ_{ij} 's, θ_{AVE} , although it does depend on the nuisance parameters $(\mu_{i0} - \mu_{0j})/\sigma$. They also derived the equation for obtaining the critical values (Hung et al. 1993, Eq. (3.4)). When the variance is known, that is, for the case of $v = \infty$, the authors provided the critical values of the standardized statistic, $\sqrt{n}T_A$, for all rectangular designs with $K \leq 5$ and $N \leq 5$.

Under the strong assumption that the dose responses μ_{ij} are increasing in i and j , Hellmich and Lehmacher (2005) considered a problem of identifying all the minimum effective doses in a combination drug study. They implemented the closed testing principle of Marcus, Peritz, and Gabriel (1976) and used the likelihood ratio and multiple contrast tests for testing each individual hypothesis.

Buchheister and Lehmacher (2006) proposed a number of closed testing procedures for identifying the set of all combinations superior to each component. One approach corresponds to constructing a family of intersection-unions of null hypotheses. This family consists of a large number of hypotheses; for example, in the 2×2 case there are 15 hypotheses to be tested. Another approach requires some adjustment of the α -level, but it does involve testing of simpler hypotheses.

3. Procedure for Identifying the Minimum Efficacious Doses

3.1 Testing Scheme

First, for a given design, all theoretically possible MeD-sets must be identified. In the $K \times N$ case, there are $\binom{K+N}{K}$ different possible MeD-sets including the empty set. For any p , $1 \leq p \leq \min\{K, N\}$, the number of possible different MeD-sets of cardinality p , can be shown to be given by the recursive formula

$$q(p, K, N) = \sum_{i=1}^{K-p+1} \sum_{j=p}^N q(p-1, K-i, j-1)$$

for $2 \leq p \leq \min\{K, N\}$ with $q(1, K, N) = K \cdot N$.

For example, in the 2×2 case, there are six MeD-sets, namely, the empty set, four single-element sets, $\{(1, 1)\}, \{(1, 2)\}, \{(2, 1)\}, \{(2, 2)\}$, and one two-element set, $\{(1, 2), (2, 1)\}$. In the 2×3 case, the preceding formulas result in nine different non-empty MeD-sets: $\{(1, 1)\}, \{(1, 2)\}, \{(1, 3)\}, \{(2, 1)\}, \{(2, 2)\}, \{(2, 3)\}, \{(1, 2), (2, 1)\}, \{(1, 3), (2, 1)\}$ and $\{(1, 3), (2, 2)\}$.

For simplicity, we present our procedure for identifying the MeD-set for the 2×2 and 2×3 cases. Our procedure can be easily established to handle higher dimensional cases, as required, although higher dimensions become increasingly more notationally complex.

After identifying all possible MeD-sets, we next construct the hypothesis family. In the 2×2 case, the family of null hypotheses is

$H_0^{(4)}: \theta_{11}=\theta_{12}=\theta_{21}=\theta_{22}=0$, $H_0^{(3)}: \theta_{11}=\theta_{12}=\theta_{21}=0$, $H_0^{(2.1)}: \theta_{11}=\theta_{12}=0$, $H_0^{(2.2)}: \theta_{11}=\theta_{21}=0$ and $H_0^{(1)}: \theta_{11}=0$

In the 2×3 case, the set of all null hypotheses is given by the following hypothesis family:

$$\begin{aligned} &H_0^{(6)}: \theta_{11}=\theta_{12}=\theta_{13}=\theta_{21}=\theta_{22}=\theta_{23}=0, \\ &H_0^{(5)}: \theta_{11}=\theta_{12}=\theta_{13}=\theta_{21}=\theta_{22}=0 \\ &H_0^{(4.1)}: \theta_{11}=\theta_{12}=\theta_{13}=\theta_{21}=0, \\ &H_0^{(4.2)}: \theta_{11}=\theta_{12}=\theta_{21}=\theta_{22}=0, \\ &H_0^{(3.1)}: \theta_{11}=\theta_{12}=\theta_{13}=0, \\ &H_0^{(3.2)}: \theta_{11}=\theta_{12}=\theta_{21}=0, \\ &H_0^{(2.1)}: \theta_{11}=\theta_{12}=0, \\ &H_0^{(2.2)}: \theta_{11}=\theta_{21}=0, \text{ and} \\ &H_0^{(1)}: \theta_{11}=0. \end{aligned}$$

Note that in our notation, each hypothesis is assigned to a certain level given by the number of gains in the hypothesis (denoted by the upper index) and that there are several hypotheses which may belong to the same level (e.g., the two hypotheses of level 4 are indexed (4.1) and (4.2)).

Each corresponding alternative hypothesis states that there is at least one efficacious combination among the combinations considered in the null hypothesis. For example, the alternative to $H_0^{(3.1)}$ is $H_a^{(3.1)}: \theta_{1j} \geq 0$ for all $j = 1, 2, 3$ and $\theta_{1j} > 0$ for some $j = 1, 2, 3$. A hypothesis family constructed in such a way is closed under intersection (see Marcus et al. 1976).

There is a one-to-one correspondence between all logically consistent possible sets of true and false null hypotheses and the population MeD-set. For example, if

$H_0^{(6)}, H_0^{(5)}, H_0^{(4.1)}$, and $H_0^{(3.1)}$ are false and the other null hypotheses are all true, then the population MeD-set is given by $\{(1, 3)\}$. Similarly, given the MeD-set $\{(1, 3)\}$ one can conclude that the hypotheses $H_0^{(6)}, H_0^{(5)}, H_0^{(4.1)}$, and $H_0^{(3.1)}$ must be false and the rest of the null hypotheses must be true.

After the family of hypotheses is constructed, we test the hypotheses in a step-down manner under the closed testing principle starting from the highest level. This guarantees the strong control of the FWE. The order for the step-down testing is given by a tree-like partial order on the set of hypotheses, where the partial order of hypotheses is defined by the set containment of the null hypotheses. When the testing is complete we are able to estimate the MeD-set based on the pattern of the resulting rejected and accepted hypotheses.

Using Figure 1 we illustrate the testing scheme for the “decision tree” for the 2×3 case (see Figure 2 for the “decision tree” for the 2×2 case). First, $H_0^{(6)}$ is tested. If it is accepted, then we stop the testing and estimate the MeD-set as the empty set. If $H_0^{(6)}$ is rejected, then $H_0^{(5)}$ is tested. If $H_0^{(5)}$ is accepted, then we stop testing and estimate the MeD-set as $\{(2, 3)\}$. Otherwise, we reject $H_0^{(5)}$ and test both hypotheses of the fourth level, and so on, unless each hypothesis

of a certain level is either tested and accepted, or not tested. A hypothesis is not tested if it is implied by a tested and accepted hypothesis or by a hypothesis which is not tested. For example, if all of the hypotheses up to and including $H_0^{(3,1)}$ were tested and rejected and $H_0^{(3,2)}$ was tested and accepted, then one should stop testing and estimate the MeD-set as $\{(1, 3), (2, 2)\}$. Table 1 summarizes the results of such a testing with the estimated MeD-sets. Similarly, a table corresponding to the 2×2 case can be constructed.

The proposed approach provides estimators that include all possible MeD-sets, but there are some decision sequences that lead to ambiguities in terms of the estimated MeD-set. For example, an ambiguity happens if we test and reject both $H_0^{(6)}$ and $H_0^{(5)}$ but then test and accept $H_0^{(4,1)}$ and $H_0^{(4,2)}$. Since $H_0^{(5)}$ cannot be false while both $H_0^{(4,1)}$ and $H_0^{(4,2)}$ are true, there is a logical inconsistency and thus there is no population MeD-set that would correspond to that sets of true and false hypotheses. Therefore, we do not provide an estimated MeD-set in this case. It is not within our current framework to decide how to estimate the MeD-set in the case of ambiguities. In an actual trial, one would prespecify the choices of MeD's for the possible ambiguous outcomes and these considerations would depend on whether there are safety concerns and one wants to possibly overestimate or underestimate the true MeD-set. In our simulations, reported in Section 5, we do report the simulation probability of ambiguities occurring.

3.2 Testing of an Individual Hypothesis: Critical Values for Nonrectangular Designs

The hypotheses constructed earlier correspond to two kinds of designs, “rectangular” and “nonrectangular.” Let us fix two numbers, K_0 and N_0 , such that $1 \leq K_0 \leq K$ and $1 \leq N_0 \leq N$. Then by a rectangular design we mean the set of combinations $G = \{(i, j), i = 1, \dots, K_0, j = 1, \dots, N_0\}$, where K_0 and N_0 denote the number of active doses of drug A and drug B currently under consideration, respectively. In the 2×3 case, any hypothesis except for $H_0^{(5)}, H_0^{(4,1)}$, and $H_0^{(3,2)}$ corresponds to such a design. The second kind of design is a nonrectangular design of the following form. Let $1 < K_0 \leq K$, $1 < N_0 \leq N$ and $1 \leq N_{K_0-1} \leq \dots \leq N_1 \leq N_0$ with $N_{K_0} < N_1$. Then the nonrectangular design is given by

$$I = \{(i, j = j(i)), j = 1, \dots, N_i, i = 1, \dots, K_0\}, \quad (2)$$

where N_i denotes the number of combinations with drug A taken at dose level i , $i = 1, \dots, K_0$. The nonrectangular design I is obtained by excluding some specified high dose combinations from the original grid of points G . For example, in the 2×3 case, one form of the nonrectangular design is given by $\{(1, 1), (1, 2), (1, 3), (2, 1), (2, 2)\}$, which corresponds to $H_0^{(5)}$.

In order to test an individual hypothesis corresponding to the rectangular design, we use the AVE test given in (1). For all hypotheses that correspond to nonrectangular designs we need to introduce a generalization of the AVE test and obtain the additional critical values.

Assume that I is given, so that K_0 and N_i 's, $i = 1, \dots, K_0$, $C_I = \sum_{i=1}^{K_0} N_i$ (the cardinality of the set I) and $P = K_0 + N_1$ (the total number of single drug doses) are specified. Next, $\hat{\sigma}^2$ is the pooled estimator of σ^2 based on the entire dataset. Then the generalization of AVE test statistic is given by

$$T_A = \sum_{(i,j) \in I} \widehat{\theta}_{ij} / (\widehat{\sigma} C_I). \quad (3)$$

Our procedure involves multiple testing and hence, several AVE test statistics for rectangular and nonrectangular designs ought to be constructed. Let us note that all of these statistics use a common estimate for variance and this estimate is based on the entire dataset.

The corresponding generalization of Hung et al. (1993) distributional results to nonrectangular designs is given by following theorem. The main steps of the proof are outlined in Appendix A.

Theorem 1—Define $G(c;I) = \int_0^\infty \int_{-\infty}^\infty \Phi(-\sqrt{nC_I}c\omega - tx)\varphi(x)dx dQ(\omega)$, where $\Phi(\cdot)$ is the distribution function of the standard normal random variable; $\varphi(\cdot)$ is its density function; $Q(\cdot)$ is the distribution function of $\hat{\sigma}/\sigma$. Then for a nonrectangular design I of the form (2), the α -level critical value of the AVE test statistic (3) is given as the solution c to the equation $G(c;$

$t_{\max}) = \alpha$, where $t_{\max} = \max \sqrt{\sum_{s=1}^P m_s^2 / C_I}$ with $m_s, s = 1, \dots, P$ nonnegative integers subject to constraints:

$$\begin{aligned} m_s &\leq N_s, s=1, \dots, K_0; m_s \leq C_{K_0}, \\ s &= K_0+1, \dots, P \text{ and } \sum_{s=1}^P m_s = C_I, \end{aligned} \quad (4)$$

where C_{K_0} denotes the cardinality of $\{(i, s - K_0) : (i, s - K_0) \in I\}$.

In order to use the preceding theorem, one needs first to solve the optimization problem of

maximizing the value of $t = \sqrt{\sum_{s=1}^P m_s^2 / C_I}$, subject to specific constraints on the values of $m_s, s = 1, \dots, P$. Appendix B illustrates the solution to this problem for some forms of the nonrectangular design. A similar approach can be used to solve this problem for any other given nonrectangular design. Table 2 provides the critical values for all nonrectangular designs for the standardized AVE test in the 2×2 and 2×3 cases for $v = \infty$, that is, for the case of known variance.

4. Antihypertensive Drugs Example

Hung et al. (1993) provided the observed diastolic blood pressure mean reductions for a 3×4 ($K = 2$ and $N = 3$ in our notation) factorial clinical trial, which is conducted to evaluate the effectiveness of the combination of two active antihypertensive drugs, A and B. The sample size of each group is 25. They noted that the observed blood pressure reductions appeared to be approximately normally distributed, with a pooled estimate for the true variance of 42.

Hung et al. (1993) tested the global null hypothesis of no positive gains versus the alternative of at least one positive gain, that is, the hypothesis $H_0^{(6)}$ at level of significance $\alpha = 0.05$. For this dataset, the estimated gains are $\hat{\theta}_{11} = 4$, $\hat{\theta}_{12} = 2$, $\hat{\theta}_{13} = 3$, $\hat{\theta}_{21} = \hat{\theta}_{22} = 1$, and $\hat{\theta}_{23} = 2$, and $\theta_{\text{AVE}} = 2.17$, where the corresponding critical value is $1.34(\sqrt{42}/\sqrt{25}) = 1.74$. Hence, Hung et al. (1993) concluded that there is at least one combination superior to its components.

The first step of our procedure, testing of $H_0^{(6)}$, is identical to testing the global hypothesis of no positive gains. Hence, in the case of $\alpha = 0.05$ the preceding results apply. Since $H_0^{(6)}$ is rejected we test the next hypothesis in the “decision tree,” and so forth. For completeness, Table 3 provides the test statistics and the critical values corresponding to each hypothesis in the “decision tree.”

For these data, the results of our procedure to estimate the MeD-set are different depending on the level of significance. If we test at $\alpha = 0.01$, then we accept $H_0^{(6)}$ and conclude that there are no minimum efficacious combinations, that is, the MeD-set is empty. Testing at $\alpha = 0.05$ results in rejection of all the hypotheses down to and including $H_0^{(2,1)}$ and acceptance of $H_0^{(2,2)}$. Hence, by Table 1, the estimated MeD-set is given by $\{(1, 2)\}$. Testing at $\alpha = 0.10$ results in rejection of all null hypotheses, so that the MeD-set is estimated as $\{(1, 1)\}$.

5. Simulation Studies

In this section we provide the details of our simulation studies for the 2×3 case. Throughout, the significance level is $\alpha = 0.05$, $\sigma = 1$ and the common sample size, $n = 30$, is assumed per group. The critical values we use are given either by Hung et al. (1993) or Table 2, divided by $\sqrt{30}$, depending on which is appropriate.

To illustrate the relationship between the overall power and gain averages we consider $\delta = \mathbf{0}$, where $\delta_{ij} = \mu_{i0} - \mu_{0j}$, so that only settings involving $\mu_{i0} = \mu_{0j}$ with $i = 1, 2, j = 1, 2, 3$, are considered. Without loss of generality, all single drug mean responses, μ_{i0} and μ_{0j} , are taken to be zero, that is, $\mu_{ij} = \theta_{ij}$.

A single simulation run is conducted as follows. For each specified value of the global gain average $\theta_{AVE} = 0, 0.4, 0.8, 1.2$, where $\theta_{AVE} = \sum_{i=1}^2 \sum_{j=1}^3 \theta_{ij}/6$, all possible configurations of the MeD-sets are considered. Since there are nine different MeD-sets that can be constructed for each of the positive global gain average, there are 28 simulation configurations in total.

The performance of our procedure depends not only on the true global average θ_{AVE} but also on all partial averages. Each partial average corresponds to a certain null hypothesis and is defined as an average of gains specified in the null hypothesis. For example, the partial gain average corresponding to $H_0^{(2,1)}$ is given by $(\theta_{11} + \theta_{12})/2$. 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 gain average, so it is impossible to construct such a study. 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 combinations. For example, if $\theta_{AVE} = 0.4$ and the MeD-set is $\{(1, 2)\}$, then $\theta_{11} = \theta_{21} = 0$ and $\theta_{12} = \theta_{13} = \theta_{22} = \theta_{23} = 0.6$.

Next, for the specified μ_{ij} 's, the group sample means $\bar{y}_{ij}$'s are generated, such that $\bar{y}_{ij}$'s are independent and $\bar{y}_{ij} \sim N(\mu_{ij}, 1/30)$. Then, our proposed procedure is applied to these data. For each simulation configuration, the result of the procedure is noted and the preceding steps are replicated 100,000 times, leading to the maximum standard error of 0.002 for the performance of characteristics of interest.

We use the following definitions. *Family-wise Error rate* (FWE) is defined as 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. And *Lack of Power* (LOP) is the probability of accepting at least one false null hypothesis.

In the single drug setting, Tamhane, Hochberg, and Dunnett (1996) used several estimated measures to assess the performance of their closed testing procedures. Hellmich and Lehman (2005) used some modifications of these notions for the combination drug setting. Similar to those authors, we obtain the estimated measures by using all available information about dose effectiveness provided by the true and estimated MeD-sets.

In order to simulate the FWE, Power, and LOP, we consider the following proportions. The proportion of replications for which at least one nonefficacious combination is estimated as an MeD corresponds to the estimated FWE of the procedure. The proportion of cases when 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 LOP is estimated by the proportion of all cases when a combination larger than the true MeD 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.

Cases that result in ambiguity are considered separately in detail. Each such case contributes to either FWE or LOP, so it is included in the calculation of estimated FWE or LOP. In addition, we consider the proportion of replications when the procedure results in ambiguity to estimate the probability of ambiguity, denoted by p_{amb} . For example, suppose the true MeD-set is given by $\{(2, 3)\}$, so that all null hypotheses except for $H_0^{(6)}$ are true. If the procedure rejects both $H_0^{(6)}$ and $H_0^{(5)}$ and then accepts both $H_0^{(4,1)}$ and $H_0^{(4,2)}$, then an ambiguity happens. In this case the true null hypothesis $H_0^{(5)}$ is rejected incorrectly. This case corresponds to the “underestimating” of the MeD (2, 3) and should be included in the calculation of the FWE.

Table 4 summarizes the results of the simulation studies for each specified value of global gain average and a particular configuration of the MeD-set. The table contains the estimated FWE, Power, LOP, p_{amb} , and the average power. In the case of the empty population MeD-set, the LOP is not defined, because all combinations are known to be nonefficacious.

Based on these results, we can draw the following conclusions. First of all, the simulations confirm the strong control of the FWE. Also they confirm that the power of the procedure depends on the value of the global gain average. In the case of the empty MeD-set (zero gain average), the estimated power is greater than 0.99. The average power is estimated as 0.246, 0.638, and 0.934 for $\theta_{\text{AVE}} = 0.4$, $\theta_{\text{AVE}} = 0.8$, and $\theta_{\text{AVE}} = 1.2$, respectively. Hence, the quality of the procedure improves with increasing positive global gain average.

Next, among possible nonempty true MeD-sets, the set containing the highest combination (the MeD-set is given by $\{(2, 3)\}$) is correctly identified with the highest probability. The “diagonal” configurations of the MeD-set are correctly identified with the lowest probabilities.

Finally, when the population MeD-set is empty, the proposed procedure almost never results in ambiguity (the proportion of such cases is less than 0.1%). The average frequency of cases resulting in ambiguity is 0.128 for $\theta_{\text{AVE}} = 0.4$, 0.121 for $\theta_{\text{AVE}} = 0.8$, and 0.013 for $\theta_{\text{AVE}} = 1.2$. Among possible configurations, the “diagonal” configurations correspond to the highest average frequency of ambiguity. The probability of ambiguity decreases with increasing positive gain average.

The simulation results for the 2×2 case similarly confirm the preceding conclusions.

6. Discussion

The problem of identifying the MeD-set is quite complex. 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 our procedure is the strong control of overall error rate 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 family-wise error rate can rapidly increase. As a consequence, a procedure involving multiple testing can result in false conclusion about the drug effectiveness if the overall error rate is not controlled.

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 component is high, there are both a large number of hypotheses to be tested and many different structures of the population MeD-sets. Next, the procedure is constructed so that all possible MeD-sets can be estimated, but together with those estimates there are rejection patterns that may result in ambiguities about how to estimate the MeD-set. In this article we do not differentiate among the possible ambiguous patterns. We are in the process of conducting additional research which explores ambiguity issues in further details and preliminary results indicate that certain ambiguities can be eliminated by generalizing our testing scheme. However, typical modern biopharmaceutical combination drug studies usually do not involve more than four or six drug combinations, and for such designs our procedure is quite effective.

Appendix

Appendices

A. On Obtaining the Additional Critical Values for the Average Test

In order to obtain the distribution function of the AVE test statistic (3) and prove the theorem one can generalize the method introduced by Hung et al. (1993) for obtaining the critical values for the AVE test (1). Here we present only final results. Let I be the specified nonrectangular design of the form (2). Then it can be shown that the power function of the AVE test statistic (3) is given by

$$P(T_A > c | \theta) = \int_0^\infty \int_{R^P} \Phi\{h_0(\mathbf{u}, \mathbf{v}, \omega, \delta, \theta)\} \cdot \prod_{i=1}^{K_0} \varphi(u_i) \prod_{j=1}^{N_1} \varphi(v_j) d\mathbf{u} d\mathbf{v} dQ(\omega),$$

where $\Phi(\cdot)$ is the standard normal distribution function; $\varphi(\cdot)$ is the standard normal density function; $Q(\cdot)$ is the distribution function of $\hat{\sigma}/\sigma$, that is, the distribution function of $(\chi_v^2/v)^{1/2}$, where χ_v^2 denotes a chi-square random variable with $v = (K+1)(N+1)(n-1)$ degrees of freedom; $\mathbf{u}$ and $\mathbf{v}$ are vectors with components u_i 's and v_j 's, respectively, for $i = 1, \dots, K_0$ and $j = 1, \dots, N_1$; δ and θ are vectors with components δ_{ij} 's and θ_{ij} 's, respectively, for $(i, j) \in I$, that is, $j = 1, \dots, N_i$, $i = 1, \dots, K_0$; $d\mathbf{u} = du_1 \dots du_{K_0}$, $d\mathbf{v} = dv_1 \dots dv_{N_1}$, $P = K_0 + N_1$, $\theta_{\text{AVE}} = \sum_{(i,j) \in I} \theta_{ij} / C_I$ and

$$h_0(\mathbf{u}, \mathbf{v}, \omega, \delta, \theta) = -\sqrt{nC_I}(c\omega - \theta_{\text{AVE}}/\sigma) - \sum_{(i,j) \in I} (u_i + v_j - \sqrt{n}|\delta_{ij}| + |u_i - v_j + \sqrt{n}\delta_{ij}|) / (2\sqrt{C_I}).$$

This power function increases in θ_{AVE} for any given δ and any fixed number c . Let $P_{\theta=0}$ denote the power function evaluated at $\theta = \mathbf{0}$, then it can be shown that

$$\text{Significance level} = \max \left\{ \lim_{|\delta_{ij}| \rightarrow \infty} P_{\theta=0}; (i, j) \in I \right\}.$$

Next, $\lim_{|\delta_{ij}| \rightarrow \infty} P_{\theta=0} = \int_0^\infty \int_{R^P} \Phi(q) \left\{ \prod_{i=1}^{K_0} \varphi(u_i) \prod_{j=1}^{N_1} \varphi(v_j) \right\} d\mathbf{u} d\mathbf{v} dQ(\omega)$, where
 $q = -\sqrt{nC_I}c\omega - \sum_{s=1}^P m_s z_s / \sqrt{C_I}$; $\pi_{ij} = 1$ if $\delta_{ij} \rightarrow \infty$, $\pi_{ij} = 0$ if $\delta_{ij} \rightarrow -\infty$; $m_s = \sum_{j:(s,j) \in I} \pi_{sj}$ and $z_s = u_s - K_0$ if $s = K_0 + 1, \dots, P$.

Let $x = \sum_{s=1}^P m_s z_s / \sqrt{\sum_{s=1}^P m_s^2}$ then $x \sim N(0, 1)$ and $\lim_{|\delta_{ij}| \rightarrow \infty} P_{\theta=0}$ can be written as

$$\lim_{|\delta_{ij}| \rightarrow \infty} P_{\theta=0} = \int_0^\infty \int_{-\infty}^\infty \Phi\left(-\sqrt{nC_I}c\omega - x \sqrt{\sum_{s=1}^P m_s^2 / C_I}\right) \varphi(x) dx dQ(\omega).$$

Finally, let $G(c; t) = \int_0^\infty \int_{-\infty}^\infty \Phi(-\sqrt{nC_I}c\omega - tx) \varphi(x) dx dQ(\omega)$, where $t = \sqrt{\sum_{s=1}^P m_s^2 / C_I} > 0$. Since $G(t)$ increases in t the significance level can be obtained as the value of $G(c; t)$ evaluated at the largest possible t , subject to the constraints of (4).

B. Obtaining the Critical Values for Nonrectangular Designs for 2×2 and 2×3 Cases

Example 1

For the nonrectangular design of the form $I_0 = G - \{(K_0, N_0)\}$, the maximum value of

$t = \sqrt{\sum_{s=1}^P m_s^2 / C_I}$ subject to the constraints (4) equals $t_{\max} = \sqrt{(ab^2 - 2b + 1)/(K_0 N_0 - 1)}$, where $a = \min\{K_0, N_0\}$ and $b = \max\{K_0, N_0\}$. Indeed, for such a design, the constraints (4) reduce to $m_s \leq N_0$, $s = 1, \dots, K_0 - 1$; $m_{K_0} \leq N_0 - 1$; $m_s \leq K_0$, $s = K_0 + 1, \dots, P - 1$; $m_P \leq K_0 - 1$ and $\sum_{s=1}^P m_s = K_0 N_0 - 1$, where $P = K_0 + N_1$.

Next, suppose that $\max\{K_0, N_0\} = K_0$. Since t attains its maximum value when it is evaluated at the highest possible values of m_s 's, we set $N_0 - 1$ numbers $m_s = K_0$ for $s = K_0 + 1, \dots, P - 1$ which add up to $K_0 N_0 - K_0 < K_0 N_0 - 1$ for $K_0 > 1$. In order to satisfy $\sum_{s=1}^P m_s = K_0 N_0 - 1$ we set $m_P = K_0 - 1$ and $m_s = 0$, $s = 1, \dots, K_0$. Hence, $t_{\max} = \sqrt{(N_0 K_0^2 - 2K_0 + 1)/(K_0 N_0 - 1)}$.

Similarly, one can consider a case of $\max\{K_0, N_0\} = N_0$ and then combine two results to get $t_{\max} = \sqrt{(ab^2 - 2b + 1)/(K_0 N_0 - 1)}$.

Hence, when the variance is known, the critical values of desired level of significance can be obtained by solving the following equation for c :

$$\alpha = \int_{-\infty}^\infty \Phi\left[-\sqrt{n(K_0 N_0 - 1)}c - \sqrt{(ab^2 - 2b + 1)/(K_0 N_0 - 1)}u\right] \varphi(u) du.$$

Example 2

Consider $K_0 = 2$, $N_0 = 3$, $N_2 = 1$ and $N_1 = 3$, that is, $I = \{(1, 1), (1, 2), (1, 3), (2, 1)\}$. Then the constraints (4) become $m_1 \leq 3$, $m_2 \leq 1$, $m_3 \leq 2$, $m_4 \leq 1$, $m_5 \leq 1$ and $\sum_{s=1}^5 m_s = 4$. We set $m_1 = 3$, $m_2 = 1$ and $m_3 = m_4 = m_5 = 0$. So $t_{\max} = \sqrt{2.5}$ and hence, in the case of known variance, the critical values can be obtained from $\alpha = \int_{-\infty}^{\infty} \Phi[-\sqrt{4nc} - \sqrt{2.5}u] \varphi(u) du$

Acknowledgments

The research of Allan Sampson was supported by NIMH Grant 5P50 MH045156-17. We wish to thank both referees, the Associate Editor and the Editor for their insightful comments and suggestions that greatly improved this article.

References

- Buchheister, B.; Lehmacher, W. Multiple Testing Procedures for Identifying Desirable Dose Combinations in Bifactorial Designs; GMS Medical Informatics, Biometry and Epidemiology. 2006. p. 1-11. Doc 07 Available online at <http://www.egms.de/pdf/journals/mibe/2006-2/mibe000026.pdf>
- Hellmich M, Lehmacher W. Closure Procedures for Monotone Bi-factorial Dose-Response Designs. Biometrics 2005;61:269–276. [PubMed: 15737103]
- Hung HM, Chi JY, Lipicky RJ. Testing for the Existence of a Desirable Dose Combination. Biometrics 1993;49:85–94. [PubMed: 8513112]
- Marcus R, Peritz E, Gabriel KR. On Closed Testing Procedures with Special Reference to Ordered Analysis of Variance. Biometrika 1976;63:655–660.
- Tamhane AC, Hochberg Y, Dunnett CW. Multiple Test Procedures for Dose Finding. Biometrics 1996;52:21–37. [PubMed: 8934584]
- U.S. Food and Drug Administrations policy. Fixed-Combination Prescription Drugs for Humans. Code of Federal Regulations, Title 21. 2006. (21 CFR 300.50). Available online at <http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?fr=300.50>

Figure 1.
“Decision tree” for the 2×3 design.

Figure 2.
“Decision tree” for the 2×2 design.

Table 1

Outcomes of the procedure for the 2×2 case.

Estimated McD-set	$H_0^{(6)}$	$H_0^{(5)}$	$H_0^{(4,1)}$	$H_0^{(4,2)}$	$H_0^{(3,1)}$	$H_0^{(3,2)}$	$H_0^{(2,1)}$	$H_0^{(2,2)}$	$H_0^{(1)}$
Empty	ACC	NT	NT	NT	NT	NT	NT	NT	NT
{(2, 3)}	REJ	ACC	NT	NT	NT	NT	NT	NT	NT
{(2, 2)}	REJ	REJ	ACC	REJ	NT	NT	NT	NT	NT
{(1, 3)}	REJ	REJ	REJ	ACC	REJ	NT	NT	NT	NT
{(2, 1)}	REJ	REJ	REJ	REJ	ACC	REJ	NT	REJ	NT
{(2, 2), (1, 3)}	REJ	REJ	REJ	REJ	REJ	ACC	NT	NT	NT
{(1, 2)}	REJ	REJ	REJ	REJ	REJ	REJ	REJ	ACC	NT
{(2, 1), (1, 3)}	REJ	REJ	REJ	REJ	REJ	REJ	ACC	REJ	NT
{(2, 1), (1, 2)}	REJ	REJ	REJ	REJ	REJ	REJ	REJ	REJ	ACC
{(1, 1)}	REJ	REJ	REJ	REJ	REJ	REJ	REJ	REJ	REJ
Ambiguity	REJ	REJ	ACC	ACC	NT	NT	NT	NT	NT
Ambiguity	REJ	REJ	REJ	ACC	ACC	NT	NT	NT	NT
Ambiguity	REJ	REJ	REJ	REJ	ACC	ACC	NT	NT	NT
Ambiguity	REJ	REJ	REJ	REJ	ACC	REJ	NT	ACC	NT
Empty	REJ	REJ	REJ	REJ	REJ	REJ	ACC	ACC	NT

NOTE: ACC = Accepted, REJ = Rejected, NT = Not Tested.

Table 2Critical values for Standardized AVE test for nonrectangular designs when variance is known, that is, $v = \infty$.

Designs with $K_0 = 2, N_0 \geq 2$	Graphical representation	Significance level		
		0.10	0.05	0.01
$N_1 = 2, N_2 = 1$	$\begin{pmatrix} * & * \\ * & \square \end{pmatrix}$	1.21	1.55	2.19
$N_1 = 3, N_2 = 1$	$\begin{pmatrix} * & * & * \\ * & \square & \square \end{pmatrix}$	1.20	1.54	2.18
$N_1 = 3, N_2 = 2$	$\begin{pmatrix} * & * & * \\ * & * & \square \end{pmatrix}$	1.09	1.40	1.97

NOTE: $\otimes$ = omitted combination.

Table 3

Gain averages and corresponding critical values for Hung et al's (1993) antihypertensive drugs example.

Hypothesis	Gain average	Critical value		
		$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
$H_0^{(6)}$	2.17	1.36	1.74	2.46
$H_0^{(5)}$	2.20	1.41	1.81	2.55
$H_0^{(4.1)}$	2.50	1.56	2.00	2.83
$H_0^{(4.2)}$	2.00	1.44	1.84	2.61
$H_0^{(3.1)}$	3.00	1.92	2.46	3.49
$H_0^{(3.2)}$	2.33	1.57	2.01	2.84
$H_0^{(2.1)}$	3.00	2.03	2.61	3.69
$H_0^{(2.2)}$	2.50	2.03	2.61	3.69
$H_0^{(1)}$	4.00	2.35	3.02	4.26

NOTE: The critical values for the rectangular and nonrectangular designs are the ones given by Hung et al. (1993) and Table 2, respectively, times $\sqrt{42}/\sqrt{25}$.

Table 4

Simulation results in the 2×3 case.

Measure	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)
$\theta_{AVE} = 0$										
Power	0.999	0.670	0.165	0.151	0.075	0.041	0.079	0.022	0.025	0.147
LOP	—	0.328	0.833	0.845	0.924	0.955	0.918	0.975	0.968	0.853
FWE	0.000	0.002	0.002	0.004	0.001	0.004	0.004	0.004	0.008	0.000
P_{amb}	0.000	0.000	0.006	0.020	0.013	0.256	0.164	0.296	0.288	0.112
$\theta_{AVE} = 0.8$										
Power	0.998	0.998	0.832	0.826	0.553	0.466	0.526	0.317	0.364	0.859
LOP	0.000	0.000	0.166	0.170	0.446	0.529	0.470	0.679	0.624	0.141
FWE	0.002	0.002	0.002	0.004	0.001	0.005	0.004	0.004	0.012	0.000
P_{amb}	0.000	0.000	0.001	0.004	0.002	0.321	0.036	0.381	0.338	0.007
$\theta_{AVE} = 1.2$										
Power	0.998	0.998	0.996	0.994	0.937	0.916	0.914	0.805	0.847	0.997
LOP	0.000	0.000	0.003	0.003	0.062	0.079	0.082	0.191	0.141	0.003
FWE	0.002	0.002	0.002	0.003	0.001	0.004	0.004	0.004	0.012	0.000
P_{amb}	0.000	0.000	0.000	0.001	0.000	0.028	0.001	0.059	0.028	0.000
Average Power	0.999	0.889	0.664	0.657	0.523	0.474	0.506	0.381	0.412	0.668

NOTE: LOP = Lack of Power, FWE = Family-wise Error rate, p_{amb} = Probability of Ambiguity; in the case of $\theta_{AVE} = 0$, the LOP is not defined. For non-empty MeD-sets the Average Power is calculated as the average of power for $\theta_{AVE} = 0.4, 0.8$, and 1.2 .