Published in final edited form as:

Pharm Stat. 2012 ; 11(3): 250–257. doi:10.1002/pst.542.

Expected loss functions as additional measures to assess performance of multiple testing procedures for combination drug dose finding

Julia N. Soualakova^{a,*} and Allan R. Sampson^b

^aDepartment of Statistics, University of Nebraska–Lincoln, Lincoln, NE, USA

^bDepartment of Statistics, University of Pittsburgh, Pittsburgh, PA, USA

Abstract

There are several measures that are commonly used to assess performance of a multiple testing procedure (MTP). These measures include power, overall error rate (family-wise error rate), and lack of power. In settings where the MTP is used to estimate a parameter, for example, the minimum effective dose, bias is of interest. In some studies, the parameter has a set-like structure, and thus, bias is not well defined. Nevertheless, the accuracy of estimation is one of the essential features of an MTP in such a context. In this paper, we propose several measures based on the expected values of loss functions that resemble bias. These measures are constructed to be useful in combination drug dose response studies when the target is to identify all minimum efficacious drug combinations. One of the proposed measures allows for assigning different penalties for incorrectly overestimating and underestimating a true minimum efficacious combination. Several simple examples are considered to illustrate the proposed loss functions. Then, the expected values of these loss functions are used in a simulation study to identify the best procedure among several methods used to select the minimum efficacious combinations, where the measures take into account the investigator's preferences about possibly overestimating and/or underestimating a true minimum efficacious combination. The ideas presented in this paper can be generalized to construct measures that resemble bias in other settings. These measures can serve as an essential tool to assess performance of several methods for identifying set-like parameters in terms of accuracy of estimation.

Keywords

average test; dose response; maximum test; minimum effective dose; stepwise procedure

1. INTRODUCTION

There are a number of statistical measures that are commonly used to assess performance of a multiple testing procedure (MTP) for dose finding and, in particular, for detecting the minimum effective dose (MED). These measures include family-wise error rate (FWE),

power, and lack of power (LOP) [1]. 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 LOP is defined as the probability of accepting at least one false null hypothesis. In addition to these measures, in consideration with the MED, bias is used to provide essential knowledge regarding how far the estimator is expected to be from the parameter. Therefore, bias is vital when several procedures are being compared for the MED in terms of expected accuracy of estimation.

Depending on the context of the problem, the definitions of these measures can be stated more explicitly in terms of the parameters and their estimates. And thus, these measures can be easily assessed via simulation. For example, if the problem of identifying the MED of a single drug with nondecreasing mean responses is considered, more specific definitions of the performance measures in terms of the true and estimated MED are used in a simulation study [2, 3]. FWE is estimated by the proportion of cases where a lower dose than the true MED is estimated as the MED, power is estimated by the proportion of cases where the true MED is estimated correctly, and LOP is estimated by the proportion of replicates where a higher dose than the true MED is declared to be the MED. To compute the bias, the one small issue to deal with is how to define the 'estimated' MED when, in fact, no MED is identified by the MTP among active doses, $1, 2, \dots, K$. In such a case, to evaluate the bias, the MED is estimated as $K + 1$, which is the smallest dose of these unexamined doses that, in theory, could have been an MED.

Although the bias serves as a very useful measure to compare several MTPs for dose finding when dealing with a single drug, it is not well defined in combination drug settings, where the parameter of interest is a set of beneficial combinations. In this paper, we provide an overview of some problems presented in the literature that illustrate these difficulties. Then, we further discuss problems associated with defining bias in these cases. To handle the concept of bias for combination drug settings, we introduce particular loss functions and their expected values.

Consider a $(K + 1) \times (N + 1)$ factorial trial, where two drugs, A and B, are combined each at several doses. Suppose that subjects are randomly allocated, n per group, to receive a certain combination (i, j) , where i and j denote the respective dose levels of drugs A and B, $i = 0, 1, \dots, K$ and $j = 0, 1, \dots, N$, and 0 denotes placebo. Let μ_{ij} denote the population mean response for combination (i, j) . Then, to measure the gain in the combination relative to the individual drug components [4], gain is defined as $\theta_{ij} = \min\{\mu_{ij} - \mu_{i0}, \mu_{ij} - \mu_{0j}\}$, $i = 1, \dots, K$, and $j = 1, \dots, N$. If $\theta_{ij} > 0$, then the combination (i, j) is called superior (to each component) [4]. 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 $\theta_{ij} \geq 0$. Consider a case where at least one of the two investigated drug components is known to be effective at the considered dose levels. In such a case, any combination with $\theta_{ij} > 0$ is effective and superior and is called efficacious [5, 6]. A minimum efficacious combination (MeD) is defined [5, 6] as an efficacious combination, such that decreasing any component dose leads to a nonefficacious compound, that is, a combination (i, j) is an MeD if $\theta_{ij} > 0$ and $\theta_{rs} = \theta_{is} = \theta_{rj} = 0$ for all $r < i$ and $s < j$.

Although in a single drug case, the MED if it exists is unique, there may be several minimum efficacious drug combinations. Therefore, the notion of the minimum efficacious dose set (MeD-set) as a collection of all MeDs was introduced [5]. For example, in the 3×3 case, there are five possible MeD-sets in addition to the empty set; these sets are illustrated in Table I.

In the problem of identifying the MeD-set, the set of gain parameters allows the determination of whether a certain combination is considered to be an MeD or not. The primary interest is to estimate the MeD-set, which is viewed as the parameter. Because neither the parameter nor any estimator is real valued, there are difficulties associated with a statistical definition of bias associated with such a set-like parameter and its estimator.

It is mentioned in the literature [7] that notion of bias can be extended to the combination drug case but no bias-like measures are presented. In this paper, we propose several expected loss functions as generalizations of bias and illustrate how MTPs can be compared in terms of the constructed measures. These methods can be useful not only in statistical methodology related to drug efficacy studies but also in bioinformatics and other fields where parameters of interest have set-like structures and bias is not well defined.

In Section 2, we review the literature about statistical methods for identifying the MeD-set. Section 3 outlines some properties of the MeD-set, which are then used to construct the proposed expected loss function measures in Section 4. Section 5 describes a simulation study conducted to evaluate performance of several MTPs in terms of the expected loss functions. A brief discussion is given in Section 6.

2. BACKGROUND

The MTPs to identify the MeD-set [5, 6] are discussed for the 3×3 factorial designs. For such designs, the family of null hypotheses is given by

$$H_0 = \{ \mu_{11} = \mu_{12} = \mu_{13} = \mu_{21} = \mu_{22} = \mu_{23} = \mu_{31} = \mu_{32} = \mu_{33} \}$$

and $H_{0i} = \{ \mu_{11} = \mu_{12} = \mu_{13} = \mu_{21} = \mu_{22} = \mu_{23} = \mu_{31} = \mu_{32} = \mu_{33} = 0 \}$, where each corresponding alternative hypothesis states that there is at least one positive gain among the gains specified in the null hypothesis as zero-gains. Each proposed MTP requires testing of individual hypotheses in a specific order. First, we review the testing order utilized by the generalized average test procedure (GAVEP) [5] and the generalized maximum test procedure (GMAXP) [6]. Then, the individual hypothesis testing details for GAVEP and GMAXP are addressed. Another procedure based on local maximum tests (LoMAXP) [6, 8] is also reviewed.

The hypotheses are each tested at level α , starting from H_0 . The testing order is specified by the closed testing principle [9]. Thus, a hypothesis is tested if and only if all hypotheses implying it were tested and rejected. In the 3×3 case, the implication relationships are quite simple: H_{01} implies H_{02} implies both H_{03} and H_{04} , and finally, each of the latter two hypotheses implies H_{05} . Therefore, under the closed testing principle, a hypothesis

is tested if and only if all other hypotheses are tested and rejected. Table I summarizes the results of such a testing strategy with the corresponding estimated MeD-sets. Note that the set of all possible population MeD-sets is the same as the set of possible estimators.

The hypothesis testing assumes that y_{ijt} is independent and $y_{ijt} \sim N(\mu_{ij}, \sigma^2)$, $i = 0, 1, \dots, K$, $j = 0, 1, \dots, N$, and $t = 1, 2, \dots, n$, where y_{ijt} is the response of the t th subject in the (i, j) th dose combination group. The pooled estimator, $\hat{\sigma}^2$, of σ^2 is given by

$$\hat{\sigma}^2 = \frac{1}{\nu} \sum_{i=0}^K \sum_{j=0}^N \sum_{t=1}^n (y_{ijt} - \bar{y}_{ij})^2, \quad (1)$$

where $\bar{y}_{ij}$ is the sample mean response for the (i, j) th group and $\nu = (K+1)(N+1)(n-1)$.

To test each hypothesis, the generalized average and generalized maximum test is used, respectively, for GAVEP and GMAXP. The tests require first evaluating 'standardized' gains [4–6, 8] as

$$T_{ij} = \frac{\bar{y}_{ij} - \max(\bar{y}_{i0}, \bar{y}_{0j})}{\hat{\sigma}}, \quad (1)$$

where $\theta_{ij} = \bar{y}_{ij} - \max(\bar{y}_{i0}, \bar{y}_{0j})$, $i = 1, \dots, K$, and $j = 1, \dots, N$. Then for each hypothesis, the generalized average test statistic is constructed as the average of T_{ij} s over the set of combinations that are specified as zero-gain combinations by the null hypothesis. Similarly,

the generalized maximum test statistic is given by the maximum of T_{ij} s, where the maximum is evaluated with respect to the set of combinations that are specified as zero-gain combinations by the null hypothesis. For example, for testing H_{11} , the generalized average test statistic is given by $(T_{11} + T_{12})/2$ and the generalized maximum test statistic is given by $\max\{T_{11}, T_{12}\}$.

The aforementioned (generalized) average and maximum tests were originally proposed to test an overall hypothesis for global superiority [4]. The key distributional properties of both tests were described in detail and the critical values for selected rectangular designs were provided [4]. Additional critical values for nonrectangular designs were also obtained for the generalized average [5] and maximum [6] tests.

The LoMAXP [8] is slightly different from the GAVEP and GMAXP as it concerns the testing of hypotheses H_{ij} versus $H_{i'j'}$, where $i = 1, 2, \dots, K$ and $j = 1, 2, \dots, N$. The test statistics T_{ij} s given in (1) are ordered from smallest to largest, and the corresponding hypotheses are tested in a step-up manner. The critical value for each test is the same α -level critical value corresponding to the generalized maximum test for the $(K+1) \times (N+1)$ case [4]. The indices of rejected and accepted hypotheses are then used to determine the MeD-set. This approach was originally proposed to identify the set of all combinations superior to each component [8]. Then, applications of LoMAXP to estimate the MeD-set were discussed [6].

The primary advantage of GAVEP, GMAXP, and LoMAXP is strong control of the overall error rate [5, 6, 8]. This feature is especially beneficial when many combinations are studied simultaneously as each test can be performed at the same significance level. On the basis of the results of a simulation study, the GMAXP was recommended because for a majority of the MeD-set configurations, it was shown to dominate the other two approaches [6]. There were cases identified where the GAVEP performed the best in terms of the power [6]. Nevertheless, in general, the LoMAXP was shown to dominate the GAVEP [6]. Moreover, because the GAVEP could result in ambiguity, the LoMAXP was recommended over it, especially in the cases where no information regarding the possible configuration of the MeD-set is available [6].

3. MINIMUM EFFICACIOUS DOSE SET

In this section, we outline some properties of the MeD-set that will be used to construct the proper loss functions. These properties depend on characteristics of an indicator gain-matrix.

Let Θ denote the $K \times N$ gain matrix with elements θ_{ij} . Then, the indicator gain-matrix, $\Lambda = \{\lambda_{ij}\}$, is defined as a $K \times N$ matrix with entries of 0 and 1 only, where 0-entries correspond to zero-gain combinations, $\theta_{ij} = 0$, and 1-entries correspond to positive-gain combinations, $\theta_{ij} > 0$. For example, Θ_0 and Λ_0 illustrate the possible gain-matrix and the corresponding indicator gain-matrix, respectively.

There is a property of isotonicity, which is not assumed in the paper but is worth mentioning here. The formal definition is as follows: Indicator gain-matrix is said to be isotonic with respect to the matrix partial order [10], $\prec$, if $(i, j) \prec (i', j')$ implies that $\lambda_{ij} \leq \lambda_{i'j'}$ for all i and j , where $i = 1, 2, \dots, K$ and $j = 1, 2, \dots, N$. Isotonicity of a gain-matrix is defined similarly. Simply stated, under isotonicity of gain indicators, efficaciousness of a certain combination implies efficaciousness of all larger (with respect to $\prec$) combinations.

Given a nonempty MeD-set, the indicator gain-matrix can be used to uniquely partition the $K \cdot N$ combinations into three components: a set, M^- , of all non-MeDs that are not larger, with respect to $\prec$, than the MeDs (these combinations are non-efficacious); the MeD-set, M ; and the set M^+ of all remaining combinations (the latter combinations are larger than the MeDs, with respect to $\prec$). Note that if the true MeD-set is empty then all combinations are allocated to M^- . For example, if in the 3×3 case the MeD-set is given by $M = \{(1, 2)\}$, then $M^- = \{(1, 1), (2, 1)\}$ and $M^+ = \{(2, 2)\}$. For the estimated MeD-set $\hat{M}$, we similarly define $\hat{M}^-$ and $\hat{M}^+$.

Note that if the indicator gain-matrix is isotonic then M^+ is a set of all efficacious combinations that are not MeDs. In addition, although isotonicity of the gain matrix implies isotonicity of the indicator gain-matrix, the converse is not true, the latter being illustrated via Θ_0 and Λ_0 given previously.

4. ON EXPECTED VALUES OF LOSS FUNCTIONS

In this section, we discuss how loss functions can be constructed so that their expected values can serve as analogs of bias to measure the accuracy of estimation of the MeD-set. Let the population MeD-set and the MeD-set estimated by a procedure be denoted by $M = \{(i, j) : (i, j) \in I \times J\}$ and $\hat{M} = \{(i, j) : (i, j) \in \hat{I} \times \hat{J}\}$, respectively, where $I \times J$ and $\hat{I} \times \hat{J}$ are some nonempty subsets of indices of $\{1, 2, \dots, K\} \times \{1, 2, \dots, N\}$. In a case where $M \cap \hat{M}$ is empty, let it be represented by $\{(K+1, N+1)\}$. Although, this set is not a valid MeD-set in the $(K+1) \times (N+1)$ case, it is logical to use such a representation to define the distance and is analogous to what is performed in the single drug MED estimation. Note that there are other possible representations of an empty set that can be incorporated. For simplicity, we use notation c_{ij} to denote a combination (i, j) . To define a loss function $L(M, \hat{M})$, we require the key features of any loss function $0 \leq L(M, \hat{M}) < \infty$, where $L(M, \hat{M}) = 0$ if and only if $\hat{M} = M$.

We propose several definitions of the loss function based on the distances between the dose combinations that belong to M and the dose combinations that belong to $\hat{M}$. We represent the sets M and $\hat{M}$ uniquely by the $K \times N$ matrices, where the entries of 1 correspond to the MeDs and the entries of 0 correspond to the other combinations. First, two definitions of the distance measures are provided. Both of these measures are used to define the loss functions.

Consider three arbitrary combinations $a_1 = (p_1, q_1)$, $a_2 = (p_2, q_2)$, and $a^* = (p^*, q^*)$, $p_1, p_2, p^* \in \{1, 2, \dots, K\}$, and $q_1, q_2, q^* \in \{1, 2, \dots, N\}$, and an arbitrary set of combinations $A = \{(p_i, q_i), 1 \leq i \leq \min\{K, N\}\}$. Then the distance between combinations a_1 and a_2 , denoted by $d(a_1, a_2)$, can be defined as $d(a_1, a_2) = |p_1 - p_2| + |q_1 - q_2|$. Similarly, the distance between a

dose combination a^* and the set A can be defined as

Definition 1

Let $M_1 \equiv \hat{M} - (M \cap \hat{M})$ and $M_2 \equiv M - (M \cap \hat{M})$. Then, the distance loss function, L_1 , is

defined as

In the case when $\hat{M}$ is not identical to M , each element of $\hat{M}$ that does not belong to M increases the value of the first sum, and each element of M that does not belong to $\hat{M}$ increases the value of the second sum in the expression for L_1 . Thus, this distance loss function accounts for all combinations being 'misclassified' to $\hat{M}^-$, $\hat{M}^+$, or $\hat{M}^\pm$. Table II presents the values of the distance loss function for all configurations of the true and estimated MeD-sets in the 3×3 case. Example 1 demonstrates calculations of some of these values.

Example 1

To illustrate the distance loss function, we consider the 3×3 design with the true MeD-set $M = \{(1, 2)\}$ and two estimated sets, $\hat{M}_1 = \{(1, 2), (2, 1)\}$ and $\hat{M}_2 = \{(2, 1)\}$. Then, $L_1(M, \hat{M}_1) = 2$ and $L_1(M, \hat{M}_2) = 4$. Thus, $\hat{M}_1$ results in smaller distance loss function than $\hat{M}_2$ does.

Although the expectation of the distance loss function can serve as a valid measure to compare several procedures, the distance loss function has a drawback of not differentiating between a set of definitely nonefficacious combinations, M^- , and a set of possibly efficacious combinations, M^+ , as is illustrated in Example 2.

Example 2

Consider the 3×3 design. Given $M = \{(1, 2)\}$, the estimated sets $M_1^{\hat{}} = \{(1, 1)\}$ and $M_2^{\hat{}} = \{(2, 2)\}$ both correspond to the distance loss functions of 2. Nevertheless, the first estimator provides a definitely nonefficacious combination as a MeD, whereas the second estimator provides an efficacious combination as a MeD (if the indicator gain-matrix is isotonic) or a combination with unknown efficaciousness (if the indicator gain-matrix is of general structure). So, it would be more logical to differentiate between these cases of estimated sets so that the corresponding loss functions are not equal.

More flexible loss functions can be constructed in the case when additional information about the combination efficaciousness is taken into account. Using the previous definitions of M^- , M , M^+ , $M^{\hat{-}}$, $M^{\hat{}}$, and $M^{\hat{+}}$, we can allocate each combination (i, j) , $i = 1, 2, \dots, K$, and $j = 1, 2, \dots, N$ to one and only one set based on its true and estimated efficaciousness. Table III summarizes all possible dose allocations with the corresponding conclusions whether or not a statistical procedure makes an estimation error. Entries 'No Error' correspond to the correct allocation. There are six types of errors. We say that a type A error occurs when a combination $(i, j) \in M \cap M^{\hat{-}}$ (the true MeD is underestimated), a type B error occurs if $(i, j) \in M \cap M^{\hat{+}}$ (the true MeD is overestimated), a type C error corresponds to the case where $(i, j) \in M^{\hat{-}} \cap M^{\hat{+}}$ (a nonefficacious combination is estimated as an MeD), and a type D error occurs if $(i, j) \in M^{\hat{-}} \cap M^{\hat{+}}$ (a nonefficacious combination is estimated as a combination larger than a true MeD). Finally, if $(i, j) \in M^+ \cap M^{\hat{-}}$ or $(i, j) \in M^+ \cap M^{\hat{+}}$, then the true MeD is overestimated, and these cases correspond to error types E and F, respectively.

The classification provides a better understanding of the desirable properties of a loss function. Clearly, the different types of errors could be weighted depending on the researcher's preference to impact the loss function. The following definition of the loss function allows assigning different weights w_A , w_B , w_C , w_D , w_E , and w_F to A, B, C, D, E, and F error types, respectively, and 0 weight to 'No Error' case. Let w_{ij} denote the weight corresponding to the error type to which the (i, j) th dose combination is assigned, so that $w_{ij} \in \{0, w_A, w_B, w_C, w_D, w_E, w_F\}$. In general, these weights can be chosen to be positive with larger values corresponding to less desirable error types.

Definition 2

Define the loss function $L_2(M, M^{\hat{}})$ by

In the succeeding paragraphs, we illustrate one of the possible assignments of the weights and calculate the corresponding loss function L_2 . Suppose that a dose response study deals with the drugs such that the larger 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 III, there are three cases that deal with overestimating the true MeD (and result in B, E, or F errors) and three cases that correspond to underestimating the MeD (and result in A, C, or D errors). In accordance with this, we choose the following weights for the error types $w_A = w_C = 1$, $w_D = 1.5$, $w_B = w_E = 2$, and $w_F = 2.5$.

Example 3

To illustrate how the loss function L_2 can be calculated based on these error values, we consider the set-up of Example 2, where M_1 underestimates the true MeD whereas M_2 overestimates it. Let Λ , Λ_1 , and Λ_2 be the indicator gain-matrices corresponding to the true and the two estimated MeD-sets, M_1 and M_2 , respectively. These matrices are given by

where, in general settings, ‘?’ corresponds to a combination that can be as efficacious as nonefficacious. If isotonicity is assumed, then ‘?’ can be replaced by 1.

Let us first calculate the loss function corresponding to M_2 . On the basis of Λ and Λ_2 , each dose combination is classified as follows: nonefficacious combinations (1,1) and (2,1) are correctly estimated as nonefficacious ones, that is, $\{(1, 1), (2, 1)\} \in M^- \cap M^-$; an MeD (1,2) is incorrectly estimated as nonefficacious, that is, $(1, 2) \in M \cap M^-$; and finally, a combination (2,2), which is larger than an MeD is estimated as an MeD, that is, $(2, 2) \in M^+ \cap M^+$. Thus, 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, $L_2(M, M_2) = 4$.

Next, let us calculate the loss function $L_2(M, M_1)$. Based on Λ and Λ_1 , $(1, 1) \in M^- \cap M^-$, $(1, 2) \in M \cap M^+$, $(2, 1) \in M^- \cap M^+$ and $(2, 2) \in M^+ \cap M^+$, 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 $L_2(M, M_1) = 3.5$. So, M_1 is preferred in terms of the loss function L_2 . This concludes the example.

Although isotonicity of an indicator gain-matrix is not assumed in the definition of L_2 , the loss function L_2 is especially useful for isotonic structures as it allows assigning different penalties for overestimating and underestimating the true MeDs.

In the general case, the true and estimated MeD-sets identify only nonefficaciousness of combinations that are not larger than the MeDs (smaller than the MeDs or noncomparable to the MeDs, with respect to $\prec$). Therefore, if we let $U = M \cup M^+$ and $\hat{U} = M \cup M^+$, we can assign the error values corresponding to each type of misclassification as follows. If a combination $(i, j) \in M^- \cap \hat{U}$, then at least one true MeD is underestimated, and thus we assign $w_{ij} = w_u > 0$. If $(i, j) \in U \cap M^-$, then at least one true MeD is overestimated so that we assign $w_{ij} = w_o > 0$. All other allocations of a combination (i, j) can be viewed as the correct ones with the corresponding allocation errors being zero. This logic then can be used to define the following new loss function L_3 as a special case of L_2 .

Definition 3

For each drug combination (i, j) , $i = 1, \dots, K$, $j = 1, \dots, N$, with the assigned value of error w_{ij} , where $w_{ij} \in \{0, w_o, w_u\}$, $w_o \geq 0$, $w_u \geq 0$, the allocation loss function is defined by

Table IV presents the values of the allocation loss function in terms of weights w_o and w_u for all configurations of the true and estimated MeD-sets in the 3×3 case. Example 4 illustrates these calculations for a specific case of w_o and w_u .

Example 4

Consider the set-up of Example 2 with error values fixed at $w_o = w_u = 1$. Let us first calculate $L_3(M, \hat{M}_2)$. Because $(1, 2) \in U \cap \hat{M}$ is the only misclassified combination, $L_3(M, \hat{M}_2) = w_o = 1$. Similarly, for $\hat{M}_1$, we have $\{(1, 1), (2, 1)\} \in \hat{M} \cap U$, so $L_3(M, \hat{M}_1) = 2w_u = 2$. Thus, $\hat{M}_2$ is preferred over $\hat{M}_1$.

Examples 2, 3, and 4 illustrate that the proposed loss functions can lead to different conclusions in terms of preferred estimates. In Example 2, the two estimated sets resulted in the same distance loss functions, whereas the first estimated set resulted in a smaller loss function in Example 3, and in a larger allocation loss function in Example 4. The choice among the loss functions depends on the primary criterion for comparison. Although the distance function is the simplest one, it does not have a flexibility to assign different penalties for overestimating and underestimating a true MeD. The second loss function, given in Definition 2, is much more flexible and takes into account all possible estimation errors and allows assigning different penalties to different types of errors. The allocation loss function permits (partially) differentiating among cases where incorrect decisions regarding combination efficaciousness were made and is relatively simple.

We have considered three definitions of the loss functions. We propose to use the expected values of these loss functions as appropriate measures for assessing accuracy of estimation. For each loss function, $L_i(M, \hat{M})$, $i = 1, 2, 3$, the corresponding expected loss function is defined as

where $P(\hat{M})$ denotes the probability of estimating the true MeD-set M by $\hat{M}$. The measures $\lambda_i(M, \hat{M})$, $i = 1, 2, 3$, then can be used as analogs of bias to make comparisons among several procedures.

5. SIMULATION STUDY

In this section, we outline the design and results of our simulation study conducted to assess performance of three estimation procedures mentioned in Section 2 in terms of the proposed measures $\lambda_1(M, \hat{M})$ and $\lambda_3(M, \hat{M})$. The design is similar to the one considered in the literature [6], where the strong control of FWE was confirmed for GAVEP, GMAXP, and LoMAXP and a number of results in terms of power of these procedures were stated. The main goal of this simulation study is to differentiate among the procedures in the case where several procedures were shown to perform relatively well in terms of the power. In these cases, where the procedures are comparable, the choice among the procedures in terms of accuracy of estimation is the most essential. Thus, the simulation configurations are chosen accordingly.

We consider the 3×3 case and evaluate the expected distance and allocation loss functions, $\lambda_1(M, \hat{M})$ and $\lambda_3(M, \hat{M})$, which are presented in Section 4. To specify the values for misclassification errors for the allocation loss function, L_3 , we consider two cases: a case where it is more desirable to overestimate a true MeD than to underestimate it (Case 1) and a case where it is more desirable to underestimate a true MeD than to overestimate it (Case 2). The weights are set to be $w_o = 1$ and $w_u = 5$ for Case 1 and $w_o = 5$ and $w_u = 1$ for Case 2. Thus, in both cases, there are penalties for misclassification of a combination, but in Case 1, the penalty for overestimating is $w_o = 1$, which is much smaller than the penalty for underestimating, $w_u = 5$, and in Case 2, these penalties are reversed. To see whether different values of weights highly affect the results, we also consider similar configurations with $w_o = 2$ and $w_u = 4$ for Case 1 and $w_o = 4$ and $w_u = 2$ for Case 2.

To estimate the expected values of the loss functions for each given true MeD-set, the proportion of replications corresponding to each estimated MeD-set is multiplied by its misclassification error given in Tables II (for L_1) and IV (for L_3) and then summed to produce a single value, so that the true probabilities are estimated by the observed ones.

Throughout the study, the significance level α is fixed at 0.05, $\sigma^2 = 1$ and the common group sample size, $n = 30$, is assumed. The component drug mean responses are fixed at $\mu_{01} = 0.2$, $\mu_{02} = \mu_{10} = 0.4$, and $\mu_{20} = 0.8$. The mean response corresponding to placebo is assumed to be 0. Thus, both component drugs are effective at the considered doses. Then, for each given MeD-set, the remaining means are chosen so that the gain-matrix is isotonic with equal maximum gains, θ , of 0.0, 0.8, and 0.9. For example, for the MeD-set $\{(1, 2)\}$ and $\theta = 0.8$, the remaining means are given by $\mu_{11} = 0.4$, $\mu_{12} = 1.2$, $\mu_{21} = 0.8$, and $\mu_{22} = 1.6$, so that the gains are $\theta_{11} = \theta_{21} = 0$ and $\theta_{12} = \theta_{22} = 0.8$. Then, all possible configurations of the MeD-set are considered for each value of maximum gain, where the empty MeD-set corresponds to a case with all gains being 0. Because there are five different MeD-sets that can be constructed for each positive maximum gain, there are 11 such simulation configurations.

Next, for the specified , the group sample means are generated, where the are independent random variables such that $\bar{y}_{ij} \sim \mathcal{N}(\mu_{ij}, 1/30)$. Then, GAVEP, GMAXP, and LoMAXP described in Section 2 are applied to these data. The result of each procedure is noted and the above steps are replicated 100 000 times for each simulation design; thus, the

maximum standard error of estimation is 0.002. Next, observed power is calculated as the proportion of replications (out of 100 000) where the true MeD-set is estimated correctly. In a case of empty MeD-set, the simulation power is estimated by a proportion of replications where the MeD-set is correctly estimated as empty. To calculate the observed values of the expected loss functions, the true probability $P(\mathcal{M})$ is estimated by the proportion of replications where the MeD-set $\mathcal{M}$ is estimated by $\hat{\mathcal{M}}$.

Table V summarizes the results in terms of power and expected loss functions for each specified value of θ and a particular configuration of the MeD-set. Because the different weighting strategies corresponding to Cases 1 ($w_o = 1$, $w_u = 5$; $w_o = 2$, $w_u = 4$) and 2 ($w_o = 5$, $w_u = 1$; $w_o = 4$, $w_u = 2$) result in similar patterns, only some results for the second weighting strategy are presented.

On the basis of all results, we can draw the following conclusions. First, the power performance is addressed. When the true MeD-set is empty or is given by $\{(1, 1)\}$, the GAVEP dominates the other procedures in terms of power. When the true MeD-set is given by $\{(2, 2)\}$, the LoMAXP dominates both methods, and for all other configurations of the MeD-set, the GMAXP performs the best in terms of the power. Note that the observed power values for LoMAXP and GMAXP are very close when the MeD-set is $\{(1, 2)\}$ or $\{(2, 1)\}$. Thus, additional information in terms of accuracy of estimation of these methods is essential to help identify the best approach.

Next, accuracy of estimation in terms of the expected distance loss function is discussed. As illustrated in Table V, the smallest values of the expected distance loss function correspond to the most powerful (for a given configuration of the MeD-set) methods. In the cases when the MeD-set is $\{(1, 2)\}$ or $\{(2, 1)\}$, GMAXP outperforms LoMAXP.

Finally, accuracy of estimation in terms of the expected allocation function is addressed. In the majority of settings, conclusions are identical to the ones based on the power. However, if the true MeD-set is $\{(1, 2)\}$ or $\{(2, 1)\}$, then LoMAXP dominates the GMAXP in Case 1 but is inferior to GMAXP in Case 2. Thus, for such configurations of the MeD-set, the GMAXP is less likely to underestimate a true MeD than are the GAVEP and LoMAXP, whereas the LoMAXP is less likely to overestimate a true MeD than are the other two methods.

6. DISCUSSION

In this paper, we propose several measures that incorporate expectations of loss functions. These measures can serve as bias-like measures to assess accuracy of estimation of an MTP for identifying all minimum efficacious dose combinations in a combination drug efficacy study. The proposed expected loss functions utilize a notion of the distance between an element and a set of elements and allow assigning different penalties for overestimating and underestimating a true minimum efficacious combination. Thus, the proposed measures are especially useful when several procedures perform well in terms of the overall error rate and power.

The results presented in this paper can be easily used in practice. If there is no prior information regarding the most likely configuration of the MeD-set, and several procedures are expected to perform similarly in terms of the power, then the following criteria can be used to select the superior methods. If an investigator would rather possibly underestimate a minimum efficacious combination to eliminate all lower (most likely nonefficacious) combinations from future studies, then the GMAXP is recommended. If, in the opposite, the investigator would prefer to possibly overestimate a minimum efficacious combination than to underestimate it, then the LoMAXP is recommended.

We note that the proposed measures can be used directly if another definition of the minimum efficacious combination is imposed, provided that the new parameter has a set-like structure similar to the one discussed in the paper. Although the research findings with respect to the considered methods are limited to the imposed assumptions (of nonnegative gains and specific distributions) and bifactorial designs, the ideas presented in this paper can be used to construct measures that resemble bias in other settings. In particular, other loss functions, which incorporate the distance and error penalties simultaneously, can be constructed. Their expected values then can be used to assess performance of a number of multiple testing procedures for identifying set-like parameters in terms of their accuracy of estimation.

In this paper, we limited our consideration to the problem of identifying all minimum efficacious combinations; however, there are other essential problems arising in combination drug efficacy studies, for example, identifying all superior combinations [8, 11, 12], all minimum effective combinations [7], and all effective and superior combinations [13–16]. Likewise, a number of important problems related to single drug efficacy studies, for example, identifying all effective doses [17], the minimum effective dose [17, 18], and the optimal dose [17], and the corresponding statistical methods have been also recently proposed. Thus, future research can be targeted at examining benefits of applications of the proposed concepts with respect to these problems and the corresponding methods. In addition, constructing an estimate for the variance of the loss functions that can be used together with the proposed mean value will help gain more information regarding the distribution of the ‘distance’ between set-like parameter and estimate.

Acknowledgments

The work of Dr. Allan Sampson was supported in part by NIMH Grant P50-MH084053.

References

1. Hochberg, Y.; Tamhane, AC. Multiple Comparison Procedures. John Wiley and Sons; New York: 1987.
2. Tamhane AC, Hochberg Y, Dunnett CW. Multiple test procedures for dose finding. *Biometrics*. 1996; 52:21–37. [PubMed: 8934584]
3. Tamhane AC, Logan BR. Multiple test procedures for identifying the minimum effective and maximum safe doses of a drug. *Journal of the American Statistical Association*. 2002; 97:293–301.
4. Hung HM, Chi JY, Lipicky RJ. Testing for the existence of a desirable dose combination. *Biometrics*. 1993; 49:85–94. [PubMed: 8513112]

5. Soulakova JN, Sampson AR. On identifying minimum efficacious doses in combination drug trials. *Statistics in Biopharmaceutical Research*. 2009; 1:39–47. [PubMed: 19675682]
6. Soulakova JN. Comparison of several testing strategies for combination drug efficacy trials based on the closure principle. *Statistics in Medicine*. 2009; 28:260–273. [PubMed: 18991315]
7. Hellmich M, Lehmacher W. Closure procedures for monotone bi-factorial dose-response designs. *Biometrics*. 2005; 61:269–276. [PubMed: 15737103]
8. Buchheister B, Lehmacher W. Multiple testing procedures for identifying desirable dose combinations in bi-factorial designs. *GMS Medical Informatics, Biometry and Epidemiology*. 2006; 2(2):1–11. Available at <http://www.egms.de/pdf/journals/mibe/2006-2/mibe000026.pdf>.
9. Marcus R, Peritz E, Gabriel KR. On closed testing procedures with special reference to ordered analysis of variance. *Biometrika*. 1976; 63:655–660.
10. Robertson, T.; Wright, FT.; Dykstra, RL. *Order Restricted Statistical Inference*. John Wiley and Sons; New York: 1988.
11. Westfall P, Ho S-Y, Prillaman B. Properties of multiple intersection-union tests for multiple endpoints in combination therapy trials. *Journal of Biopharmaceutical Statistics*. 2001; 11(3):125–138. [PubMed: 11725927]
12. Frommolt P, Hellmich M. Resampling in multiple-dose factorial designs. *Biometrical Journal*. 2009; 51:915–931. [PubMed: 20029896]
13. Soulakova JN. Resampling-based and other multiple testing strategies with application to combination drug trials with factorial designs. *Statistical Methods in Medical Research*. 2011; 20(5):505–521. [PubMed: 20667933]
14. Soulakova JN. Comparison of gatekeeping and other testing methods for identifying superior drug combinations in bi-factorial designs with isotonic parameters. *Journal of Biopharmaceutical Statistics*. 2011; 21(4):635–649. [PubMed: 21516561]
15. Soulakova JN. General multistage gatekeeping procedures for identifying beneficial drug combinations in factorial trials with isotonic gains. *Statistics in Biopharmaceutical Research*. 2010; 2(1):33–41.
16. Soulakova JN. On identifying effective and superior drug combinations via Holm's procedure based on the Min tests. *Journal of Biopharmaceutical Statistics*. 2009; 19(2):280–291. [PubMed: 19212880]
17. Bretz F, Hsu J, Pinheiro J, Liu Y. Dose finding - a challenge in statistics. *Biometrical Journal*. 2008; 50:480–504. [PubMed: 18663758]
18. Soulakova JN, Sampson AR, Jia G, Gleser LJ. Simultaneous confidence bands for comparisons to placebo, with application to detecting the minimum effective dose. *Journal of Biopharmaceutical Statistics*. 2011; 22(1):93–108. [PubMed: 22204529]

Table I

Estimated minimum efficacious dose set as a result of multiple testing procedures [5, 6].

Estimated MeD-set
Empty	ACC	NT	NT	NT	NT
{(2, 2)}	REJ	ACC	NT	NT	NT
{(2, 1)}	REJ	REJ	ACC	REJ	NT
{(1, 2)}	REJ	REJ	REJ	ACC	NT
{(1, 2), (1, 1)}	REJ	REJ	REJ	REJ	ACC
{(1, 1)}	REJ	REJ	REJ	REJ	REJ

MeD-set, minimum efficacious dose set; ACC, accepted; REJ, rejected; NT, not tested.

Table II

Values of the distance loss function for all configurations of the true and estimated MeD-sets in the 3×3 case.

<i>M</i>	<i>M</i> ^ˆ					
	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

M and *M*^ˆ denote the true and estimated Med-sets, respectively.

Table III

Possible allocations of a single combination and the corresponding errors.

Population property of a combination	Estimated property of a combination		
	Nonefficacious combination (M^-)	MeD ($M^\wedge$)	Combination larger than MeD (M^+)
Nonefficacious combination (M^-)	No error	Error C	Error D
MeD (M)	Error B	No error	Error A
Combination larger than MeD (M^+)	Error F	Error E	No error

MeD denotes minimum efficacious combination; sets M^- , M , M^+ , $M^\wedge$, $M^\wedge$, and M^f are defined in Section 4.

Table IV
Values of the allocation loss function for all configurations of the true and estimated MeD-sets in the 3×3 case.

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

M and $\hat{M}$ denote the true and estimated Med-sets, respectively.

Table V

Simulation results for generalized average test procedure (upper entry), local maximum test procedure (middle entry), and generalized maximum test procedure (bottom entry).

Measure	Case	Minimum efficacious dose set					
		Empty	{(2, 2)}	{(2, 1)}	{(1, 2)}	{(1, 1)}	
		$\theta = 0$	$\theta = 0.8$				
Power		0.980	0.245	0.316	0.278	0.297	0.909
		0.973	0.790	0.801	0.682	0.558	0.783
		0.973	0.784	0.812	0.699	0.667	0.891
Expected distance loss function		0.085	2.985	2.262	2.285	1.135	0.239
		0.157	0.813	0.734	0.893	1.104	0.601
		0.158	0.829	0.711	0.859	0.896	0.325
Expected allocation loss function	$w_o = 1$ $w_u = 5$	0.138	0.803	0.958	0.993	0.672	0.126
		0.282	0.336	0.337	0.479	0.549	0.365
		0.285	0.387	0.390	0.563	0.509	0.199
		0.026	3.720	4.446	4.491	2.666	0.609
	$w_o = 5$ $w_u = 1$	0.058	1.005	1.386	1.847	2.614	1.810
		0.059	1.015	1.298	1.696	1.927	0.976
$\theta = 0.9$							
Power			0.301	0.406	0.367	0.409	0.962
			0.875	0.887	0.809	0.734	0.882
			0.868	0.892	0.817	0.809	0.954
Expected distance loss function			2.749	1.733	1.741	0.858	0.105
			0.471	0.367	0.462	0.592	0.317
			0.489	0.357	0.445	0.460	0.125
Expected allocation loss function	$w_o = 1$ $w_u = 5$		0.748	0.780	0.815	0.545	0.049
			0.254	0.202	0.315	0.316	0.169
			0.313	0.266	0.411	0.319	0.072
	$w_o = 5$ $w_u = 1$		3.420	3.582	3.587	2.008	0.244
			0.558	0.720	0.999	1.431	0.853

Measure	Case	Minimum efficacious dose set						
		Empty $\theta = 0$	$\theta = 0.8$					
			$\{(2, 2)\}$	$\{(2, 1)\}$	$\{(1, 2)\}$	$\{(2, 1), (1, 2)\}$	$\{(1, 1)\}$	$\{(1, 1)\}$
Expected allocation loss function	$w_0 = 2$ $w_\theta = 4$	0.107	0.569	0.662	0.891	0.950	0.364	0.364
		0.228	1.417	1.486	1.510	0.913	0.099	0.099
		0.230	0.327	0.331	0.480	0.590	0.341	0.341
	$w_o = 4$ $w_\theta = 2$	0.055	0.373	0.368	0.526	0.479	0.147	0.147
		0.114	2.759	2.881	2.905	1.651	0.196	0.196
		0.115	0.480	0.579	0.830	1.146	0.690	0.690
			0.499	0.554	0.772	0.780	0.297	0.297

θ denotes maximum gain.