POWER-ENHANCED MULTIPLE DECISION FUNCTIONS CONTROLLING FAMILY-WISE ERROR AND FALSE DISCOVERY RATES

By Edsel A. Peña and Joshua Habigery

Technical Report #2009-2
January 21, 2009

This material was based upon work partially supported by the National Science Foundation under Grant DMS-0635449 to the Statistical and Applied Mathematical Sciences Institute. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Statistical and Applied Mathematical Sciences Institute
PO Box 14006
Research Triangle Park, NC 27709-4006
www.samsi.info

Submitted to the Annals of Statistics
arXiv: math.PR/0000000

POWER-ENHANCED MULTIPLE DECISION FUNCTIONS CONTROLLING FAMILY-WISE ERROR AND FALSE DISCOVERY RATES

By Edsel A. Peña[*,‡] and Joshua Habiger[†,‡]

University of South Carolina, Columbia[‡]

Improved procedures, in terms of smaller missed discovery rates (MDR), for performing multiple hypothesis testing with control of the family-wise error rate (FWER) or the false discovery rate (FDR) are presented and studied. The improvement over existing procedures such as the Sidak procedure for FWER control and the Benjamini-Hochberg (BH) procedure for FDR control is achieved by exploiting possible differences in the power functions of the individual tests. The results signal the need to take into account the power function of individual tests and to have multiple hypothesis decision functions that are not simply limited to using the individual p-values, as is the case for example with the Sidak, Bonferonni, or BH procedures. The mathematical framework utilized is decision theoretic, and by using auxiliary randomizers, the procedures could be utilized even with discrete-type or mixed-type data, in contrast to existing p-value based procedures whose theoretical validity is contingent on the uniformity of the p-value statistic under the null hypothesis. The proposed procedures are relevant in the analysis of high-dimensional "large M, small n" data sets, whose generation and creation is being accelerated by advances in high-throughput technology, notably, but not limited to, microarray technology.

1. Introduction and Motivation. Advances in modern technology, spearheaded by the microarray, has led to the creation or generation of many data sets characterized by a large number, M, of sets of variables, with the mth set $\mathcal{S}_m$ composed of variables which pertain to characteristics of an observational unit, which for historical reasons is usually called a 'gene', and with the variables $\mathbf{Z}_m$ in $\mathcal{S}_m$ measured or observed only at a small

---

[*]E. Peña is Professor in the Department of Statistics, University of South Carolina, Columbia.

[†]J. Habiger is a PhD student in the Department of Statistics, University of South Carolina, Columbia.

AMS 2000 subject classifications: Primary 62F03; secondary 62J15

Keywords and phrases: Benjamini-Hochberg procedure, Bonferonni procedure, false discovery rate (FDR), family wise error rate (FWER), Lagrangian optimization, Neyman-Pearson most powerful test, microarray analysis, reverse martingale, missed discovery rate (MDR), multiple decision function, multiple hypotheses testing, optional sampling theorem, power function, Sidak procedure.

number of replications. The variables in each set may come in varied types such as being continuous, categorical, discrete, mixed, or even as stochastic processes, and possess an inherent data structure such as being a multi-group data, a regression-type data, or even as an event-time data with covariates and in the presence of right-censoring or truncation. Such data sets may symbolically be represented by the collection of random elements

$$\texttt{DATA} \equiv \{\mathbf{Z}_{mj} : \ j = 1, 2, \ldots, n_m; \ m = 1, 2, \ldots, M\}.$$

Later, in order to simplify notation and to introduce more abstraction and generality, the observables for gene m will simply be denoted by $X_m = \{\mathbf{Z}_{mj} : \ j = 1, 2, \ldots, n_m\}$. In (8), for example, are described four such data sets. The first is a prostate data set from (26) with $M = 6033$ genes and for each gene is associated a variable Z_{m1} indicating presence (value $= 1$) or absence (value $= 0$) of prostate cancer and a variable Z_{2m} representing a continuous response. For the mth gene the random vector $\mathbf{Z}_m = (Z_{m1}, Z_{m2})$ was observed on $n = 102$ replications, and these 102 observations were utilized to compare the diseased ($Z_{m1} = 1$) and the non-diseased ($Z_{m2} = 0$) groups with respect to the response variable Z_{m2} using a two-sample t-test. The other three data sets described in (8) were an education data set from (19) with $M = 3748$; a proteomics data set from (33) with $M = 230$ and $n = 551$; and an imaging data from (23) with $M = 15445$ and $n = 12$. In all of these data sets, there is usually a decision to be made for each gene, with the decision possibly being a choice between two competing hypotheses, or obtaining an estimate of some parameter of interest, or even predicting the value of some function of $\mathbf{Z}_{m0}$, a new observation or measurement of $\mathbf{Z}_m$.

Thus, in essence, these "large M, small n" data sets are the inputs in multiple decision problems, also called in (8) as parallel inference problems, with the most common type being that of simultaneous or multiple hypothesis testing. In the latter, for the mth gene, there is a null hypothesis H_{m0} and an alternative hypothesis H_{m1} for which a choice is to be made based on the DATA. These problems in turn have spurred considerable research activity among researchers, notably statisticians, since in performing such multiple decision-making, there is a need to be cognizant and be cautious of the *Hyde*-ian nature of multiplicity, though some other procedures, especially those with an empirical Bayes flavor (cf., (8)), exploit the *Jekyll*-ian potentials of multiplicity. In multiple hypothesis testing this entails being able to hold a tenuous balance between two competing desires: to control the rate at which correct null hypotheses are erroneously rejected, but also to maintain an ability to discover correct alternative hypotheses.

Similarly to classical single-pair hypothesis testing, an error that is committed when a correct null hypothesis is rejected is referred to as of Type I, while one that is committed when a false null hypothesis is not rejected is of Type II. There are several types of Type I error rates in the multiple testing scenario, and we refer the reader to (4) and (5) for a comprehensive discussion of a variety of these Type I error rates. The first one of interest to us is the weak family wise error rate (FWER), which is the probability of rejecting at least one null hypothesis when all the nulls are correct. The second one of interest to us is the false discovery rate (FDR), introduced by (1) in their seminal paper, which is the expected proportion of the number of false rejections of nulls relative to the number of rejections. Terminology-wise, a rejection of a null hypothesis is oftentimes called a *discovery* in multiple testing. On the other hand, the Type II error rate of interest to us will be the missed discovery rate (MDR), which is the expected proportion of the number of false non-rejections of nulls relative to the number of false null hypotheses. There are other Type II error rates in the multiple testing setting that could be considered, and these are discussed for instance in (4), (29), (3), (7), and (5). We will justify our focus on the MDR in Section 2.

Analogously to classical hypothesis testing, in multiple hypothesis testing the commission of a Type I error is considered more serious than that of a Type II error. Therefore, one framework in the development of multiple decision functions requires that one control a chosen Type I error rate at a pre-specified level, while making the MDR, or another Type II error rate, small and if possible, minimal. For example, a procedure that controls the weak FWER, under an independence assumption among the genes, is the Sidak procedure (25); while a more conservative one, but which does not require the independence condition, is the Bonferroni procedure. For control of the FDR, the procedure introduced by Benjamini and Hochberg (1), hence referred to as the BH procedure, achieves the desired control. Other works have dealt with related Type I error measures to the FDR. The papers by (9), (6), (7), and (32) discussed controlling the mFDR, an error rate which is asymptotically equivalent to the FDR (10). On the otherhand, (28) and (29) dealt with the pFDR, which is also similar to the FDR and related to the local FDR in (8), the latter having a Bayesian justification. Some more other papers, such as (24), (15), and (14), focused on the estimation of the proportion of correct null hypotheses.

Many of these multiple hypotheses testing procedures, such as the Sidak, Bonferonni, and BH procedures, rely on the set of significance or p-values of the individual tests. It should be noted that the validity of these p-value based multiple testing procedures is anchored on the technical requirement that each p-value statistic be distributed as a standard uniform when the null hypothesis is correct, a requirement which is not satisfied with non-continuous variables or with nonparametric tests such as the Mann-Whitney-Wilcoxon two-sample test. More importantly, it is not apparent whether such p-value based multiple testing procedures are utilizing, if at all, the individual power or power function of each of the M tests, since in most of these procedures, there is a common threshold that is determined, possibly in a data-dependent manner, and for those genes whose p-value is smaller than this threshold their associated null hypotheses are rejected. But what if the individual tests have different powers or power functions? Since power functions are germane for control of Type II error, is it still reasonable to impose a common threshold on all the p-values without compromising the ability to control Type II error rates?

These are the motivating questions and the prime catalyst of this paper. We will examine this issue in a decision-theoretic framework allowing for general data types and structures so results are applicable even for discrete or mixed type data and when using rank-based nonparametric tests. We will exploit the power functions of the individual test procedures to develop optimal and/or improved multiple testing procedures that control FWER or FDR. These procedures also possess smaller Type II error rates. We surmise that it is more the rule, rather than the exception, that in multiple hypotheses testing, the individual tests will have different power traits, owing to varied distributional characteristics among the X_ms, such as differing variabilities of measured or observed variables, differing effect sizes of interest, and possibly the use of different tests as dictated by the data type or structure, such as when some of the tests are t-tests, others are chi-square or analysis of variance F tests, and some others are nonparametric tests. Multiple testing procedures which rely on the usual p-value statistics and a rejection rule determined by a common threshold, as is the case with the Sidak, Bonferonni, and BH procedures, will be unable to exploit such differences in the power characteristics of the individual tests.

Papers that have utilized the power functions of the individual tests in the multiple testing situation are quite few, and the optimality constraint imposed is usually with respect to control of the expected number of false rejections. Among these papers are those by (27) which considers a simple null versus a simple alternative hypotheses for each gene, as well as those of (21) and (30). On the otherhand, (11) considered weighted p-value based procedures with control of the FDR, where the weights are chosen according to *a priori* probabilities of the null hypothesis being correct, while (34) deals with the setting where the distributions of the individual test statistics are approximately normal, or when they have differentiable density functions, cf., (22). These papers are dealing with different situations compared to the setting considered in the current paper.

In their pioneering work ushering the era of the Neyman-Pearson framework of single-pair hypothesis testing, (18) demonstrated that the most basic, and indeed the most fundamental, type of hypotheses in the single-pair hypothesis testing problem is with a simple null hypothesis and a simple alternative hypothesis. Their Fundamental Lemma which reveals the existence and uniqueness of a most powerful (MP) test function in this simple null versus simple alternative hypotheses setting opened the doors to optimal classes of test functions in more complicated settings, leading to classes of test functions which possess properties such as uniformly most powerful (UMP), UMP unbiased, or UMP invariant, and the exploitation of the monotone likelihood ratio (MLR) property of certain classes of distributions. (16) provides a comprehensive account of this Neyman-Pearson framework of hypothesis testing, a framework which dictates that in the search for optimal test functions the role of the power function is central and paramount. This framework also led to the divorce from the purely significance or p-value based approach to hypothesis testing which was then dominant during the first quarter of the 20th century.

It appears that, in a parallel manner, we are currently in the same juncture for the multiple hypothesis testing problem as almost a century ago. Many current multiple testing procedures are p-value based and do not exploit the power functions of the individual tests. It behooves to examine if better multiple testing procedures will arise by utilizing the individual power functions, in parallel to what Neyman and Pearson did in the single-pair hypothesis testing problem. This paper is a modest attempt in this direction. By considering the most basic, but we believe is also the most fundamental, setting in this multiple hypothesis testing situation, we will study multiple decision functions in the situation where for each gene, the null and alternative hypotheses are both simple. In the search for multiple decision functions this will allow as a starting point the most powerful test for each of the M pairs of hypotheses, which exists by virtue of the Neyman-Pearson Fundamental Lemma. Each of these MP test will have a power function, but as we will see, it is beneficial to look at these power functions as functions of the tests' sizes. We note that the label 'power function' may be construed as a misnomer since this is not as a function of a distributional parameter as is in the usual convention, but rather is a function of the size of the test.

We now outline the contents of this paper. In Section 2 we present the decision-theoretic framework which will serve as a platform for obtaining the multiple decision functions. This entails describing the probability models, an independence condition underlying the model, relevant loss functions, multiple decision functions, and risk functions. The Type I and Type II error rates of interest will be informed by the choice of loss functions. We also justify here the choice of the MDR as the Type II error rate of interest. Section 3 provides both a review and a re-examination of most powerful tests and p-values, in particular properties of the power function when viewed as a function of its size, which will become central in later developments. We will then utilize the results of Section 3 to find the optimal FWER-controlling procedure in Section 4. The existence will be addressed in subsection 4.1, whereas subsection 4.2 will deal with the uniqueness. Section 5 provides an explicit method for determining the optimal solution when the power functions are differentiable with respect to their sizes. Section 6 will provide a discussion concerning situations with composite hypotheses but in the presence of the monotone likelihood ratio property, and relate the theory to effect sizes. Section 7 relates the FWER-controlling optimal procedure to p-value based procedures and the distributions of the p-value statistics. Section 8 illustrates the theory for specific concrete multiple testing situations. In subsection 8.1 a normal distribution model is considered; in subsection 8.2 an exponential model is illustrated; while subsection 8.3 deals with a Bernoulli model which has the purpose of demonstrating the applicability to discrete models. In these concrete examples we also demonstrate the gain in efficiency of the optimal FWER-controlling procedure relative to the Sidak procedure. In subsection 8.4 we discuss an interesting feature of the procedure which has bearing in investing the overall FWER-size to each of the tests, with an interesting tangential application to contemporary political events!

Section 9 deals with an improved procedure which controls the FDR and possesses better Type II error rate performance than the BH procedure. The development of this new procedure is anchored on the FWER-controlling optimal procedure in the earlier sections. It will be shown that the BH procedure is a special case of the more general procedure, and it is of interest to note that martingale arguments were utilized to prove that the procedure does achieve FDR control. A consequence is a martingale-based proof of FDR-control by the BH procedure, though we note that such a martingale proof for the BH procedure has already been presented in (31), but for the more restricted setting with continuous-type data. In Section 10 we provide a modest simulation study demonstrating that the new procedure does improve on the BH procedure with respect to the MDR, at least for the normal model in this simulation. Concluding remarks are provided in Section 11.

2. Mathematical Setting. Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a basic probability space on which all random entities are defined, and $\mathcal{M} = \{1, 2, \ldots, M\}$ be an index set, where M is a known positive integer. For each $m \in \mathcal{M}$, let $X_m : (\Omega, \mathcal{F}) \to (\mathcal{X}_m, \mathcal{B}_m)$, where $\mathcal{X}_m$ is some space with associated σ-field of subsets $\mathcal{B}_m$. Form the product space $(\mathcal{X}, \mathcal{B})$ with $\mathcal{X} = \times_{m\in\mathcal{M}}\mathcal{X}_m$ and $\mathcal{B} = \sigma\left(\times_{m\in\mathcal{M}}\mathcal{B}_m\right)$. Thus, $X = (X_1, X_2, \ldots, X_M) : (\Omega, \mathcal{F}) \to (\mathcal{X}, \mathcal{B})$. The induced probability measure of X is $Q = \mathbf{P}X^{-1}$, while the (marginal) probability measure of X_m is $Q_m = PX_m^{-1}$, which is also

$$Q_m(B_m) = Q(\mathcal{X}_1 \times \ldots \times \mathcal{X}_{m-1} \times B_m \times \mathcal{X}_{m+1} \times \ldots \times \mathcal{X}_M), \forall B_m \in \mathcal{B}_m.$$

For each $m \in \mathcal{M}$, let Q_{m0} and Q_{m1} be two known probability measures on $(\mathcal{X}_m, \mathcal{B}_m)$. In conjunction with the simplication of first treating the situation of simple null hypothesis versus simple alternative hypothesis for each $m \in \mathcal{M}$, we shall assume that Q belongs to $\mathcal{Q}$, which is the collection of all probability measures on $(\mathcal{X}, \mathcal{B})$ such that the marginal probability measures Q_ms satisfy the conditions that $Q_m \in \{Q_{m0}, Q_{m1}\}$ for each $m \in \mathcal{M}$. We shall then denote by $\theta = (\theta_1, \ldots, \theta_M) : \mathcal{Q} \to \Theta \equiv \{0, 1\}^M$ with $\theta_m(Q) = I\{Q_m = Q_{m1}\}$. We shall refer to $\theta(Q)$ as the state of of the marginal probability measures of Q. Define, for each $Q \in \mathcal{Q}$, the subcollections

$$\mathcal{M}_0 \equiv \mathcal{M}_0(Q) = \{m \in \mathcal{M} : \theta_m(Q) = 0\}; \tag{2.1}$$

$$\mathcal{M}_1 \equiv \mathcal{M}_1(Q) = \{m \in \mathcal{M} : \theta_m(Q) = 1\}. \tag{2.2}$$

In this paper we shall impose the following *independence condition:* $(X_m, m \in \mathcal{M}_0(Q))$ is an independent collection of random entities. That is,

$$Q(\cap_{m\in\mathcal{M}_0(Q)}\{X_m \in B_m\}) = \prod_{m\in\mathcal{M}_0(Q)} Q_m(B_m), \forall B_m \in \mathcal{B}_m. \tag{2.3}$$

On the otherhand, the collection $(X_m, m \in \mathcal{M}_1(Q))$ need not be an independent collection, but we do assume that this collection is independent of $(X_m, m \in \mathcal{M}_0(Q))$. Two extremal subcollections of $\mathcal{Q}$ of interest later are

$$\mathcal{Q}_0 = \{Q \in \mathcal{Q} : \theta_m(Q) = 0, \forall m \in \mathcal{M}\}; \tag{2.4}$$

$$\mathcal{Q}_1 = \{Q \in \mathcal{Q} : \theta_m(Q) = 1, \forall m \in \mathcal{M}\}. \tag{2.5}$$

Observe that by virtue of the independence condition, $\mathcal{Q}_0$ is a singleton set and we shall denote by Q_0 its element. $\mathcal{Q}_1$, on the other hand, is not necessarily a singleton set.

Now, stated in its most basic form, the relevant decision problem is to determine the subcollections $\mathcal{M}_0(Q)$ and $\mathcal{M}_1(Q)$, on the basis of a realization of X. We may restate this decision problem as a multiple hypotheses testing problem where one is interested in simultaneously testing, based on a realization of X, the M pairs of hypotheses $H_{m0} : Q_m = Q_{m0}$ versus $H_{m1} : Q_m = Q_{m1}$ for $m \in \mathcal{M}$. The pairs of hypotheses could also be stated in terms of the θ-vector via $H_{m0} : \theta_m(Q) = 0$ versus $H_{m1} : \theta_m(Q) = 1$.

We shall approach this problem, however, in the more formal decision-theoretic framework with the following elements. The relevant *action space* is $\mathcal{A} = \{0,1\}^M$ with generic element $a = (a_1, a_2, \ldots, a_M)^t \in \mathcal{A}$ with the interpretation that $a_m = 0(1)$ means that H_{m0} is accepted (rejected). The relevant *parameter space* is $\mathcal{Q}$, though the effective parameter space is $\Theta = \{0,1\}^M$ with generic element $\theta = (\theta_1, \theta_2, \ldots, \theta_M)^t$. For this decision problem, we introduce several *loss functions*, $L : \mathcal{A} \times \mathcal{Q} \rightarrow \Re_+$, which are defined via:

$$L_{0k}(a,Q) = I\{a^t(1-\theta(Q)) \geq k\}, \; k = 1, 2, \ldots, M; \tag{2.6}$$

$$L_1(a,Q) = \left[\frac{a^t(1-\theta(Q))}{a^t 1}\right] I\{a^t 1 > 0\}; \tag{2.7}$$

$$L_2(a,Q) = \left[\frac{(1-a)^t \theta(Q)}{\theta(Q)^t 1}\right] I\{\theta(Q)^t 1 > 0\}. \tag{2.8}$$

The interpretations of these loss functions are as follows. The loss function $L_{0k}(a,Q)$ equals 1 if and only if at least k false discoveries are committed (with discovery being a rejection of a null hypothesis), so that in particular, when $k = 1$, $L_{01}(a,Q)$ becomes 1 if and only if at least one false discovery is committed. The loss $L_1(a,Q)$ could be interpreted as the *false discovery rate*, which was introduced by (1) in their seminal paper, since it is the ratio of the number of falsely declared discoveries and the number of declared discoveries; whereas the loss $L_2(a,Q)$ could be interpreted as the *missed discovery rate* since it is the proportion of true alternative hypotheses that were not discovered. We focus on this type of missed discovery rate since the relevant question to us is *what proportion of correct alternatives ($\theta(Q)^t 1$) were missed by using the action a?* In the literature, other types of losses, such as the rate of false negatives $[(1-a)^t\theta(Q)/(1-a)^t 1]$, have also been considered.

For this decision problem, a *nonrandomized* (multiple hypothesis testing) decision function (MHTDF) is a $\delta : (\mathcal{X}, \mathcal{B}) \rightarrow (\mathcal{A}, \sigma(\mathcal{A}))$, where $\sigma(\mathcal{A})$ could be taken to be the power set of $\mathcal{A}$. Such a decision function may be represented by

$$\delta(x) = (\delta_1(x), \delta_2(x), \ldots, \delta_M(x))^t$$

where $\delta_m(x) \in \{0,1\}$. Note that in general each δ_m could be made to depend on the full data x instead of just x_m. We shall denote by $\mathcal{D}$ the class of all nonrandomized decision functions.

More generally, we could consider *randomized* decision functions. Let us denote by $\mathcal{P}(\mathcal{A})$ the space of all probability measures over $(\mathcal{A}, \sigma(\mathcal{A}))$. A randomized decision function is a $\delta^* : (\mathcal{X}, \mathcal{B}) \to (\mathcal{P}(\mathcal{A}), \sigma(\mathcal{P}(\mathcal{A})))$, and for a realization $X = x$, an action is randomly chosen from $\mathcal{A}$ using the probability measure $\delta^*(x)(\cdot)$. We shall denote by $\mathcal{D}^*$ the space of all possible randomized decision functions, and clearly $\mathcal{D} \subset \mathcal{D}^*$ since nonrandomized decision functions are degenerate randomized decision functions. By augmenting the data X with a randomizer $U \sim U(0,1)$, which is independent of X, randomized decision functions could be made into nonrandomized decision functions. To see this, let $b_1, b_2, \ldots, b_{2^M}$ be an enumeration of the 2^M (M-tuple) elements of $\mathcal{A}$. Then $\mathcal{P}(\mathcal{A})$ may be identified with the simplex

$$\mathcal{Z} = \left\{ z = (z_1, z_2, \ldots, z_{2^M}) : \ z_j \in [0,1], \sum_{j=1}^{2^M} z_j = 1 \right\}, \tag{2.9}$$

since an element $z \in \mathcal{Z}$ represents a probability function over $\mathcal{A}$ in that z_j could be viewed as the probability of element b_j. Thus, every $\delta^*(x)$ corresponds to a $z \in \mathcal{Z}$, while every $z \in \mathcal{Z}$ is a realization of a randomized decision function. Let $\delta^*(x)$ be identified with $z \in \mathcal{Z}$, and let u be the realization of the randomizer $U \in U[0,1]$, which is generated independently of X. To choose an action from $\mathcal{A}$ according to $\delta^*(x)$ and u, action b_j is chosen if and only if $\sum_{i=1}^{j-1} z_i < u \leq \sum_{i=1}^{j} z_i$. Therefore, by utilizing the augmented data (X, U), it suffices to consider only nonrandomized decision functions. Henceforth, $\mathcal{D}$ will represent all nonrandomized decision functions $\delta(X, U)$s based on the augmented data (X, U).

For each $\delta \in \mathcal{D}$ and the different loss functions defined earlier, we have the corresponding risk functions. These are defined via:

$$R_{0k}(\delta, Q) = \mathbf{P}\left\{ \sum_{m \in \mathcal{M}} \delta(X,U)^t (1 - \theta(Q)) \geq k \right\}; \tag{2.10}$$

$$R_1(\delta, Q) = E_{(X,U)} \left\{ \frac{\delta(X,U)^t (1 - \theta(Q))}{\delta(X,U)^t 1} I\{\delta(X,U)^t 1 > 0\} \right\}; \tag{2.11}$$

$$R_2(\delta, Q) = E_{(X,U)} \left\{ \frac{(1 - \delta(X,U))^t \theta(Q)}{\theta(Q)^t 1} I\{\theta(Q)^t 1 > 0\} \right\}. \tag{2.12}$$

Associated with MHTDF $\delta = (\delta_1, \delta_2, \ldots, \delta_M)^t$ is a vector of power functions

$$\pi_\delta(Q) = (\pi_{\delta_1}(Q), \pi_{\delta_2}(Q), \ldots, \pi_{\delta_M}(Q))^t$$

where

$$\begin{aligned} \pi_{\delta_m}(Q) &= E_{(X,U)}\{\delta_m(X,U)|X\sim Q\} \\ &= \mathbf{P}\{\delta_m(X,U)=1|X\sim Q\},\ m\in\mathcal{M}. \end{aligned} \tag{2.13}$$

We may now re-express (2.12) using this power function vector via

$$R_2(\delta,Q)=\frac{(1-\pi_\delta(Q))^t\theta(Q)}{\theta(Q)^t 1}I\{\theta(Q)^t 1>0\}. \tag{2.14}$$

In terms of these risk functions, the (weak) family-wise error rate (FWER) is $\mathrm{FWER}(\delta)=R_{01}(\delta,Q_0)$, and more generally, the (weak) k-FWER is $k\text{-FWER}(\delta)=R_{0k}(\delta,Q_0)$ for $k\geq 1$. Observe that if each δ_m depends only on X_m and U, then by virtue of the independence condition,

$$\mathrm{FWER}(\delta)=1-E_U\left\{\prod_{m\in\mathcal{M}}\left[1-\mathbf{P}_{X_m\sim Q_{m0}}\{\delta_m(X_m,U)=1|U\}\right]\right\}. \tag{2.15}$$

An alternative formulation under the case $Q=Q_0$ and with the mth component δ_m^* of the randomized decision function δ^* depending only on X_m is, instead of just having one randomizer U, to have instead $U=(U_1,U_2,\ldots,U_M)$ consisting of independent and identically distributed $U[0,1]$ variables and independent of the X_ms. The mth component $\delta_m^*(X_m)(\in[0,1])$ may then be redefined via $\delta_m(X_m,U_m)=I\{U_m\leq\delta_m^*(X_m)\}$, which makes $\delta(X,U)=(\delta_m(X_m,U_m),m\in\mathcal{M})^t$ a nonrandomized decision function depending on (X,U). In this case, (2.15) could be re-expressed as

$$\mathrm{FWER}(\delta)=1-\prod_{m\in\mathcal{M}}\left[1-\mathbf{P}_{X_m\sim Q_{m0}}\{\delta_m(X_m,U_m)=1\}\right]. \tag{2.16}$$

The risk function $R_1(\delta,Q)$ will also be called the (expected) false discovery rate (FDR) of δ at Q; while the risk function $R_2(\delta,Q)$ will be referred to as the (expected) missed discovery rate (MDR) of δ at Q.

In analogy with the Neyman-Pearson framework for testing a null hypothesis versus an alternative hypothesis, the risks FWER, k-FWER, and FDR will be considered as Type I errors, whereas the risk MDR will be viewed as a Type II error. Type I errors will then be considered as more serious types of errors relative to the Type II error. As a consequence, in the search for good decision functions, the magnitude of the Type I error will usually be controlled to be no more than a pre-specified threshold, and subject to this constraint, a decision function is chosen such that the magnitude of the Type II error is small. Thus, for (weak) FWER-control, a threshold $\alpha\in(0,1)$ is specified, and a decision function δ^* is sought such that

$R_{01}(\delta, Q_0) = \text{FWER}(\delta^*) \le \alpha$, and for any other δ satisfying $\text{FWER}(\delta) \le \alpha$, $\sup_{Q\in\mathcal{Q}} R_2(\delta^*, Q) \le \sup_{Q\in\mathcal{Q}} R_2(\delta, Q)$. Observe, however, that for any $\delta \in \mathcal{D}$, and using the representation of $R_2(\delta, Q)$ in (2.14),

$$\sup_{Q\in\mathcal{Q}} R_2(\delta, Q) = \sup_{Q\in\mathcal{Q}_1} R_2(\delta, Q) = 1 - \frac{1}{M} \inf_{Q\in\mathcal{Q}_1} \sum_{m\in\mathcal{M}} \pi_{\delta_m}(Q). \tag{2.17}$$

Thus, the optimality condition on the MDR is equivalent to maximizing the sum of the powers of the individual components of the decision function. An analogous requirement exists for k-FWER control. For FDR-control, on the other hand, a threshold $q^* \in (0, 1)$ is pre-specified, and it is desired to find a decision function δ^* such that, *whatever Q is*,

$$R_1(\delta^*, Q) \le q^* \tag{2.18}$$

and such that for any other δ satisfying $R_1(\delta, Q) \le q^*$, we have that

$$\inf_{Q\in\mathcal{Q}_1} \sum_{m\in\mathcal{M}} \pi_{\delta^*_m}(Q) \ge \inf_{Q\in\mathcal{Q}_1} \sum_{m\in\mathcal{M}} \pi_{\delta_m}(Q).$$

The FDR constraint in (2.18) needs to hold true whatever the unknown Q is, and this is referred to as *strong* control, in contrast to the FWER constraint which is only required at $Q_0 \in \mathcal{Q}_0$, referred to as *weak* control.

3. Revisiting MP Tests and P-Values.

Towards the goal of finding good decision functions in line with the optimality criteria described above, we shall begin by examining the subclass of decision functions $\mathcal{D}_0 \subset \mathcal{D}$ consisting of $\delta(X, U)$ whose mth component depends only on (X_m, U_m), that is, of form $\delta_m(X_m, U_m)$. Restricted to this class, we shall determine the optimal FWER-controlling decision function, and using this optimal FWER-controlling procedure as an anchor, develop an adaptive FDR-controlling procedure which is expected to possess some optimality properties in the sense of having a smaller MDR relative to the (1) procedure. In this section, we shall therefore revisit the Neyman-Pearson most powerful test and obtain properties pertaining to its power function, but as a function of its size. Since many existing decision functions in this multiple testing setting are based on the set of significance or P-values, we also examine properties of the P-value statistic.

3.1. *Most Powerful Tests.*

For this section we first consider the case where $X : (\Omega, \mathcal{A}) \to (\mathcal{X}, \mathcal{B})$ is an observable random entity. Denote by

$Q = \mathbf{P}X^{-1}$ the probability measure of X. Consider the problem of testing, based on a realization of X, the pair of hypotheses

$$H_0(\text{Null}) : Q = Q_0 \quad \text{versus} \quad H_1(\text{Alternative}) : Q = Q_1, \tag{3.1}$$

where Q_0 and Q_1 are two probability measures on $(\mathcal{X}, \mathcal{B})$. Denote by q_0 and q_1 versions of the density functions of Q_0 and Q_1, respectively, with respect to some fixed dominating measure ν, e.g., $\nu = Q_0 + Q_1$ or it could be counting or Lebesgue measure.

A test or decision function is a measurable function

$$\delta : (\mathcal{X}, \mathcal{B}) \to ([0,1], \sigma[0,1]),$$

where $\sigma[0,1]$ is the Borel sigma-field on $[0,1]$. Given $X = x$, $\delta(x)$ is the probability of deciding in favor of H_1. The size of a test function δ is given by

$$\eta_\delta = E_{Q_0}\delta(X). \tag{3.2}$$

δ is said to be of level of significance α, where $\alpha \in [0,1]$, if $\eta_\delta \le \alpha$. The power of a test function δ of size η at $Q = Q_1$ will be denoted by

$$\pi_\delta = \pi_\delta(\eta) = \pi_\delta(Q_1; \eta) = E_{Q_1}\delta(X). \tag{3.3}$$

A test function δ^* is said to be most powerful (MP) of level η if $\eta_{\delta^*} \le \eta$ and for any other test function δ with $\eta_\delta \le \eta$, we have that $\pi_{\delta^*} \ge \pi_\delta$.

Let $L : (\mathcal{X}, \mathcal{B}) \to (\Re_+, \sigma(\Re_+))$ be a version of the likelihood ratio function so that $L(x) = q_1(x)/q_0(x)$ a.e. $[\nu]$, with the convention that $0/0 = 0$. Denote by $G_0(\cdot)$ and by $G_1(\cdot)$ the distribution functions of $L(X)$ when $\mathcal{L}(X) = Q_0$ and $\mathcal{L}(X) = Q_1$, respectively, where $\mathcal{L}(X)$ means the probability measure or law of X. For a monotone nondecreasing right-continuous function $M(\cdot)$ from $\Re$ into $\Re$, we define

$$M^{-1}(r) = \inf\{x \in \Re : \ M(x) \ge r\} \quad \text{and} \quad \Delta M(r) = M(r) - M(r-).$$

The Neyman-Pearson Fundamental Lemma ((16; 18)) states that the MP test function of level η for testing H_0 versus H_1 is given by

$$\delta^*(X;\eta) \equiv \delta^*(\eta) = I\{L(X) > c(\eta)\} + \gamma(\eta) I\{L(X) = c(\eta)\} \tag{3.4}$$

where $c(\eta) = G_0^{-1}(1-\eta)$ and $\gamma(\eta) = (G_0(c(\eta)) - (1-\eta))/\Delta G_0(c(\eta))$. Let $U \sim U(0,1)$ be independent of X. This will be called an auxiliary randomizer. We may redefine δ^* via

$$\delta^{**}(X, U; \eta) = \delta^{**}(\eta) = I\{\delta^*(X;\eta) = 1\} + I\{\delta^*(X;\eta) = \gamma(\eta); U \le \gamma(\eta)\}.$$

This test is now nonrandomized, so that with the aid of an auxiliary randomizer, the MP test, and in fact any other test, could always be made nonrandomized. The power (at $Q = Q_1$) of this MP test, either in the randomized or the nonrandomized form, is

$$\pi_{\delta^*(\eta)}(\eta) = \pi_{\delta^*(\eta)}(Q_1; \eta) = 1 - G_1(c(\eta)) + \gamma(\eta)\Delta G_1(c(\eta)). \tag{3.5}$$

We will be interested in the mapping

$$\eta \mapsto \pi_{\delta^*(\eta)}(\eta) \tag{3.6}$$

and with some abuse of terminology, still refer to this as the power function (recall that in our usual terminology the power function is viewed as a function of the parameter). We now present some properties of this function.

PROPOSITION 3.1. *The power $\pi_{\delta^*(\eta)}(\eta)$ in (3.5) of the most powerful test δ^* of size η as a function of η satisfies (i) for each $\eta \in [0,1]$, $\pi_{\delta^*(\eta)}(\eta) \geq \eta$, which implies in particular that $\pi_{\delta^*(1)}(1) = 1$, and (ii) it is concave, continuous, and nondecreasing.*

PROOF. The first result follows from the unbiasedness property of the most powerful test. To prove the concavity, suppose that it is not concave. Then there exists $\eta_1 \in [0,1]$, $\eta_2 \in [0,1]$, and $\xi \in (0,1)$ such that

$$\xi\pi_{\delta^*(\eta_1)}(\eta_1) + (1-\xi)\pi_{\delta^*(\eta_2)}(\eta_2) > \pi_{\delta^*(\xi\eta_1+(1-\xi)\eta_2)}(\xi\eta_1 + (1-\xi)\eta_2). \tag{3.7}$$

Consider the test function $\delta^{**} = \xi\delta^*(\eta_1) + (1-\xi)\delta^*(\eta_2)$. The size of this test is equal to $\xi\eta_1 + (1-\xi)\eta_2$, while its power is equal to $\xi\pi_{\delta^*(\eta_1)}(\eta_1) + (1-\xi)\pi_{\delta^*(\eta_2)}(\eta_2)$. From (3.7) the power of δ^{**} therefore exceeds the power of the MP test whose size is $\xi\eta_1 + (1-\xi)\eta_2$. Since δ^{**} has the same size as this MP test, this leads to a contradiction. Therefore the function $\eta \mapsto \pi_{\delta^*}(\eta)$ must be concave. Because of concavity, it must also be continuous. Furthermore, since $\eta \leq \pi_{\delta^*(\eta)}(\eta) \leq 1$ with $\pi_{\delta^*(1)}(1) = 1$, then it also follows by concavity that it is nondecreasing. □

We remark here that even though $\pi_{\delta^*(1)}(1) = 1$, it is not always the case that $\pi_{\delta^*(0)}(0) = 0$. For example, consider the case where $H_0 : Q_0 = U[0,1]$ while $H_1 : Q_1 = U[0,2]$. Then, with one observation X, a most powerful test of level η is $\delta^*(X;\eta) = I\{X \geq 1-\eta\}$. The power of this test is $\pi_{\delta^*(\eta)}(\eta) = [2-(1-\eta)]/2 = (1+\eta)/2$, which equals $1/2$ when $\eta = 0$.

In Proposition 3.1 we have established that the mapping $\eta \mapsto \pi_{\delta^*(\eta)}(\eta)$ is nondecreasing. In the following proposition we address the question on when it is strictly increasing. This result will be of importance in determining uniqueness of the optimal solution in the multiple hypothesis testing considered later (see Corollary 4.1).

PROPOSITION 3.2. *On the set* $\{\eta \in [0,1] : \pi_{\delta^*(\eta)}(\eta) < 1\}$, $\pi_{\delta^*(\eta)}(\eta)$ *is strictly increasing in* η.

PROOF. Since $G_0(\cdot)$ is a distribution function, then it has at most a countable number of discontinuities. Let $\{y_l\}$ represent the discontinuities of $G_0(\cdot)$. For each discontinuity y_l there is an interval $J_l = (G_0(y_l-), G_0(y_l)] \subseteq [0,1]$ such that $\forall u \in J_l$, $G_0^{-1}(u) = c_l$. Observe that the $\{J_l\}$s are disjoint and the $\{c_l\}$s are distinct. Let $J = \cup_l J_l$ and $J^c = [0,1] \setminus J$. We need to show that for every $\eta_1, \eta_2 \in [0,1)$ with $\eta_1 < \eta_2$ and $\pi_{\delta^*(\eta_2)}(\eta_2) < 1$, $\pi_{\delta^*(\eta_1)}(\eta_1) < \pi_{\delta^*(\eta_2)}(\eta_2)$.

First, consider the case where $c(\eta_1) = c(\eta_2) (= c_l)$, so that $\eta_1, \eta_2 \in J_l$ for some l. Since c_l is a jump point of $G_0(\cdot)$, then $\Delta G_0(c_l) > 0$. Suppose that $\Delta G_1(c_l) = 0$. Let $\mathcal{X}_1 = \{x \in \mathcal{X} : L(x) = c_l\}$, so that $Q_0(\mathcal{X}_1) > 0$ whereas $Q_1(\mathcal{X}_1) = 0$. Since $L(x) = q_1(x)/q_0(x)$ a.e. $[\nu]$, then $q_1(x) = 0$ a.e. $[\nu]$ on $\mathcal{X}_1$, so that $L(x) = 0$ a.e. $[\nu]$ on $\mathcal{X}_1$. But then this implies that $c_l = 0$. In this situation, the power of $\delta^*(\eta_2)$ becomes, noting that $\Delta G_1(c_l) = 0$,

$$\pi_{\delta^*(\eta_2)}(\eta_2) = 1 - G_1(c_l) = 1 - G_1(0) = 1 - \Delta G_1(0) = 1.$$

But this will contradict the condition that $\pi_{\delta^*(\eta_2)}(\eta_2) < 1$, hence we could not have $\Delta G_1(c_l) = 0$. As a consequence, since $\eta_1 < \eta_2$,

$$\begin{aligned}
\pi_{\delta^*(\eta_1)}(\eta_1) &= 1 - G_1(c_l) + \left[\frac{G_0(c_l) - (1-\eta_1)}{\Delta G_0(c_l)}\right] \Delta G_1(c_l) \\
&< 1 - G_1(c_l) + \left[\frac{G_0(c_l) - (1-\eta_2)}{\Delta G_0(c_l)}\right] \Delta G_1(c_l) = \pi_{\delta^*(\eta_2)}(\eta_2).
\end{aligned}$$

Next, consider the case where $c(\eta_1) > c(\eta_2)$. This implies that η_1, η_2 do not both belong to some J_l. Since the number of jump points of G_0 is at most countable, then there exists a η', η'' with $\eta_1 \le \eta' < \eta'' \le \eta_2$ with $c(\eta') > c(\eta'')$ and also with $c(\eta')$ and $c(\eta'')$ both G_0-continuity points. We must therefore also have $\gamma(\eta') = \gamma(\eta'') = 0$, so that

$$\pi_{\delta^*(\eta')}(\eta') = 1 - G_1(c(\eta')) \quad \text{and} \quad \pi_{\delta^*(\eta'')}(\eta'') = 1 - G_1(c(\eta'')).$$

Suppose that $\pi_{\delta^*(\eta')}(\eta') = \pi_{\delta^*(\eta'')}(\eta'')$. Then G_1 must be flat on $[c(\eta''), c(\eta')]$, while at the same time, $G_0(c(\eta'')) - G_0(c(\eta')) > 0$. Let $\mathcal{X}_2 = \{x \in \mathcal{X} : L(x) \in (c(\eta''), c(\eta'))\}$. Then we have $Q_1(\mathcal{X}_2) = 0$ and $Q_0(\mathcal{X}_2) > 0$. This implies that $q_1(x) = 0$ a.e. $[\nu]$ on $\mathcal{X}_2$, hence $L(x) = 0$ a.e. $[\nu]$ on $\mathcal{X}_2$. But then $0 \notin (c(\eta''), c(\eta'))$, implying that $\nu(\mathcal{X}_2) = 0$. This contradicts $Q_0(\mathcal{X}_2) > 0$. Therefore it is not possible to have $\pi_{\delta^*(\eta')}(\eta') = \pi_{\delta^*(\eta'')}(\eta'')$, hence we must have by nondecreasing property of $\pi_{\delta^*(\eta)}(\eta)$, that $\pi_{\delta^*(\eta')}(\eta') < \pi_{\delta^*(\eta'')}(\eta'')$. Again, by nondecreasing property, this implies that $\pi_{\delta^*(\eta_1)}(\eta_1) < \pi_{\delta^*(\eta_2)}(\eta_2)$. This completes the proof of the proposition. □

3.2. *Significance or p-Values.* Let x_0 be a realization of X. The p-value or the significance value associated with this data realization generated by the MP test is defined via

$$S(x_0) = Q_0\{L(X) \geq L(x_0)\} = 1 - G_0(L(x_0)-). \tag{3.8}$$

The $S : (\mathcal{X}, \mathcal{B}) \to ([0,1], \sigma[0,1])$ with $S(x)$ defined in (3.8) will be referred to as the p-value or significance value statistic. We shall denote by $H_0(\cdot)$ and by $H_1(\cdot)$ the distribution functions of $S(X)$ when $\mathcal{L}(X) = Q_0$ and $\mathcal{L}(X) = Q_1$, respectively.

Proposition 3.3. *For* $\forall s \in [0,1]$, $H_0(s) = 1 - G_0[G_0^{-1}(1-s)]$ *and* $H_1(s) = 1 - G_1[G_0^{-1}(1-s)]$.

Proof. For $i \in \{0,1\}$, we have $H_i(s) = Q_i\{S(X) \leq s\} = Q_i\{1 - G_0(L(X)-) \leq s\} = Q_i\{L(X) > G_0^{-1}(1-s)\} = 1 - Q_i\{L(X) \leq G_0^{-1}(1-s)\} = 1 - G_i[G_0^{-1}(1-s)]$. □

In general, we have that $G_0[G_0^{-1}(r)] \geq r$ with equality occurring whenever $G_0^{-1}(r)$ is a continuity point of $G_0(\cdot)$. Consequently, we have the following corollary.

Corollary 3.1. *If $G_0(\cdot)$ is a continuous distribution function, then $H_0(s) = s$, that is, $S(X)$ has a standard uniform distribution under $\mathcal{L}(X) = Q_0$.*

Under the condition that $G_0(\cdot)$ is a continuous distribution function, note from the definition of $\gamma(\eta)$ for the MP test that this will always equal zero. Consequently, in this situation, the power function of this MP test becomes $\pi_{\delta^*(\eta)}(\eta) = 1 - G_1[G_0^{-1}(1-\eta)]$, which is exactly $H_1(\eta)$. We state this formally as a corollary.

Corollary 3.2. *If the distribution function $G_0(\cdot)$ of the likelihood ratio statistic $L(X)$ is continuous, then the power of the Neyman-Pearson MP test of level η for testing $H_0 : Q = Q_0$ versus $H_1 : Q = Q_1$ equals the value of* *the distribution function of the p-value statistic $S(X)$ under $Q = Q_1$ when evaluated at η.*

We note that in both Corollaries 3.1 and 3.2 we needed to assume the continuity of the distribution function $G_0(\cdot)$. This will not be satisfied when dealing with discrete or mixed data. We may extend these results, however, by utilizing the so-called *randomized* p-value statistics. These are the p-value analogs of randomized tests or decision functions. This notion will be discussed in another paper, as this is not central to what we wanted to demonstrate in the current paper.

4. Optimal FWER Control. Let us now return to the decision problem formulated in Section 2. For each $m \in \mathcal{M}$, let ν_m be a chosen dominating measure in $(\mathcal{X}_m, \mathcal{B}_m)$ of Q_{m0} and Q_{m1}, which always exists since it could be taken for instance to be $\nu_m = Q_{m0} + Q_{m1}$. For $m \in \mathcal{M}$ and $j \in \{0, 1\}$, denote by $q_{mj} : (\mathcal{X}_m, \mathcal{B}_m) \to (\Re_+, \sigma(\Re_+))$ a version of the density function of Q_{mj} with respect to ν_m. Then let $L_m(x_m) = q_{m1}(x_m)/q_{m0}(x_m)$ be a version of the likelihood ratio statistic, and by $S_m(x_m)$ the p-value statistic associated with $X_m = x_m$. Denote by $\delta^*_m(\eta_m)$ the Neyman-Pearson MP test of level η_m for testing H_{m0} versus H_{m1}, and denote its power (at $Q_m = Q_{m1}$) by $\pi_{\delta^*_m(\eta_m)}(\eta_m) \equiv \pi_m(\eta_m)$.

For an MHTDF $\delta = (\delta_1, \delta_2, \ldots, \delta_M) \in \mathcal{D}_0$ with δ_m having size η_m, from (2.16), its (weak) FWER is given by

$$\text{FWER}(\delta) = 1 - \prod_{m \in \mathcal{M}} (1 - \eta_m). \tag{4.1}$$

With respect to the search for an optimal MHTDF, we are looking for a $\delta^* \in \mathcal{D}_0$ such that $\text{FWER}(\delta^*) \le \alpha$ for a pre-specified $\alpha \in (0, 1)$ and such that $R_2(\delta^*, Q_1) \le R_2(\delta, Q_1)$ for all $Q_1 \in \mathcal{Q}_1$ and for any other $\delta \in \mathcal{D}_0$ satisfying $\text{FWER}(\delta) \le \alpha$. From (2.17), $R_2(\delta, Q_1)$ is minimized if and only if $\frac{1}{M} \sum_{m \in \mathcal{M}} \pi_m(\eta_m)$ is maximized. This entails that we choose for each $m \in \mathcal{M}$ a δ_m that is most powerful, so that at each $m \in \mathcal{M}$ we need to employ the Neyman-Pearson MP test. The problem therefore reduces to choosing the appropriate sizes for each of these M MP tests.

Thus, for FWER-control, the optimization problem amounts to determining the size vector $\eta = (\eta_1, \eta_2, \ldots, \eta_M)$ that will maximize

$$\frac{1}{M} \sum_{m \in \mathcal{M}} \pi_{\delta^*_m(\eta_m)}(\eta_m) \equiv \frac{1}{M} \sum_{m \in \mathcal{M}} \pi_m(\eta_m) \tag{4.2}$$

subject to the constraint that

$$1 - \text{FWER}(\delta^*) = \prod_{m \in \mathcal{M}} (1 - \eta_m) \ge 1 - \alpha. \tag{4.3}$$

At this point we mention two possible choices for the size vector $\eta = (\eta_1, \eta_2, \dots, \eta_M)$ that satisfies the FWER constraint (4.3). The Sidak procedure (25) has these sizes all equal, with $\eta_m = 1-(1-\alpha)^{1/M},\ m \in \mathcal{M}$, which guarantees that the FWER is exactly equal to α, and this requires for its validity the assumed independence condition in (2.3). A conservative choice of η, but still assuming equal sizes, is the Bonferroni inequality-derived choice with $\eta_m = \alpha/M,\ m \in \mathcal{M}$. This choice also satisfies the FWER constraint (4.3), though equality is not achieved. However, it does not require the independence condition in (2.3).

We point out however that the optimization problem as stated is still not yet the most general problem since we restricted our search of δ^* to be in $\mathcal{D}_0$, that is, that the test function δ_m should only depend on X_m (and U_m). This amounts to having 'rectangular' rejection regions in the space $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_M$ when the test functions δ_ms are nonrandomized. The more general optimization problem is to allow the regions to be nonrectangular, equivalently to have each δ_m depend on the whole data vector $\mathbf{X} = (X_1, X_2, \dots, X_M)$. The BH procedure (1), which will be discussed later, is an example of such an MHTDF and these procedures are often referred to as adaptive, or sometimes sequential, MHTDFs, and they are usually derived by starting with the p-value statistics of the M initial test functions δ_ms, where the initial test function δ_m depends only on X_m. See (4) and (5) for a review of some of these adaptive procedures.

4.1. *Existence of Optimal Size Vector.* We establish in this section the existence of an optimal size vector for FWER control. Let

$$\mathcal{N} = [0,1]^M = \{\eta = (\eta_1, \eta_2, \dots, \eta_M) : \eta_m \in [0,1], m \in \mathcal{M}\}$$

be the size space. For $\alpha \in [0,1]$, define

$$C_\alpha = \begin{cases} \{\eta \in \mathcal{N} : \sum_{m\in\mathcal{M}} \log(1-\eta_m) \geq \log(1-\alpha)\} & \text{if} \quad \alpha < 1 \\ \mathcal{N} & \text{if} \quad \alpha = 1 \end{cases}, \tag{4.4}$$

the FWER constraint set. The following proposition provides some properties of C_α.

PROPOSITION 4.1. *The constraint set C_α has the following properties: (i) $\eta = \mathbf{0} \in C_\alpha$; (ii) $(\mathbf{0}, \alpha_m) \in C_\alpha$ for all $m \in \mathcal{M}$, where $(\mathbf{0}, \alpha_m)$ is the zero-vector but with the mth element replaced by α; and (iii) it is convex and closed.*

PROOF. The results clearly holds when $\alpha = 1$ for in this case $C_\alpha = \mathcal{N}$. For $\alpha \in [0,1)$, results (i) and (ii) are immediate, while the closedness of C_α

follows from the continuity of the logarithmic function. Let $\eta_1, \eta_2 \in C_\alpha$ with $\eta_1 \neq \eta_2$, and let $\xi \in (0,1)$. Since $\sum_{m\in\mathcal{M}} \log(1-\eta_{jm}) \geq \log(1-\alpha), j = 1,2,$ and the mapping $\eta \mapsto \log(1-\eta)$ is strictly concave, then it follows that

$$\begin{aligned} &\sum_{m\in\mathcal{M}} \log[1-(\xi\eta_{1m}+(1-\xi)\eta_{2m})] \\ &\quad > \quad \xi \sum_{m\in\mathcal{M}} \log(1-\eta_{1m}) + (1-\xi)\sum_{m\in\mathcal{M}} \log(1-\eta_{2m}) \\ &\quad \geq \quad \xi \log(1-\alpha) + (1-\xi)\log(1-\alpha) = \log(1-\alpha). \end{aligned}$$

This establishes the convexity of C_α. □

Next, we define upper sets and upper boundary sets.

DEFINITION 4.1. *The upper set of an* $\eta_0 \in \mathcal{N}$ *is* $U(\eta_0) = \{\eta \in \mathcal{N} : \eta_m \geq \eta_{0m}, \forall m \in \mathcal{M}\}$.

DEFINITION 4.2. *The upper boundary set of the constraint set* C_α *associated with an* $\alpha \in [0,1)$ *is* $UB(C_\alpha) = \{\eta \in \mathcal{N} : C_\alpha \cap U(\eta) = \{\eta\}\}$.

PROPOSITION 4.2. *For* $\alpha \in [0,1)$,

$$UB(C_\alpha) = \left\{\eta \in \mathcal{N} : \sum_{m\in\mathcal{M}} \log(1-\eta_m) = \log(1-\alpha)\right\}.$$

PROOF. Let $\eta \in UB(C_\alpha)$ so $\{\eta\} = C_\alpha \cap U(\eta)$. Suppose $\sum_{m\in\mathcal{M}} \log(1-\eta_m) > \log(1-\alpha)$. Then by continuity of the logarithm function, there exists an $\epsilon > 0$ such that $\eta + \mathbf{1}\epsilon \in \mathcal{N}$ and $\sum_{m\in\mathcal{M}} \log(1-\eta_m) > \sum_{m\in\mathcal{M}} \log[1-(\eta_m+\epsilon)] \geq \log(1-\alpha)$. Thus, $\eta+\mathbf{1}\epsilon \in C_\alpha$ and clearly $\eta+\mathbf{1}\epsilon \in U(\eta)$. Consequently, $\eta + \mathbf{1}\epsilon \in C_\alpha \cap U(\eta)$ which contradicts the fact that $\{\eta\} = C_\alpha \cap U(\eta)$. Therefore, we must have $\sum_{m\in\mathcal{M}} \log(1-\eta_m) = \log(1-\alpha)$.

On the other hand, let $\eta \in \mathcal{N}$ such that $\sum_{m\in\mathcal{M}} \log(1-\eta_m) = \log(1-\alpha)$. Then $\eta \in C_\alpha$, and since $\eta \in U(\eta)$, it follows that $\eta \in C_\alpha \cap U(\eta)$. Suppose there exists an $\eta_1 \in \mathcal{N}$ with $\eta_1 \neq \eta$ and $\eta_1 \in U(\eta)$. Then, $\eta_1 = \eta + \mathbf{\Delta}$ with $\Delta_m \geq 0$ for all $m \in \mathcal{M}$ with strict inequality for some $m \in \mathcal{M}$. Therefore, $\sum_{m\in\mathcal{M}} \log(1-\eta_{1m}) = \sum_{m\in\mathcal{M}} \log(1-\eta_m-\Delta_m) < \sum_{m\in\mathcal{M}} \log(1-\eta_m) = \log(1-\alpha)$. This implies that $\eta_1 \notin C_\alpha$. Therefore we must have $\{\eta\} = C_\alpha \cap U(\eta)$, hence $\eta \in UB(C_\alpha)$. □

Next let us define for $b \in [0,1]$ the subset of $\mathcal{N}$, denoted by $\mathcal{N}_b$, via

$$\mathcal{N}_b = \left\{\eta \in \mathcal{N} : \sum_{m\in\mathcal{M}} \pi_m(\eta_m) \geq Mb\right\}. \tag{4.5}$$

Proposition 4.3. *For $b \in [0,1]$, $\mathcal{N}_b$ satisfies (i) $\eta = \mathbf{1} \in \mathcal{N}_b$ and (ii) it is closed and convex.*

Proof. From Proposition 3.1, $\forall m \in \mathcal{M}, \pi_m(1) = 1$, hence $\mathbf{1} \in \mathcal{N}_b$. It was also established in the same proposition that $\eta \mapsto \pi_m(\eta)$ is continuous, nondecreasing, and concave. That $\mathcal{N}_b$ is closed follows from the continuity of each $\pi_m(\cdot)$. The convexity of $\mathcal{N}_b$ follows from the concavity of each of the $\pi_m(\cdot)$ analogously in the proof of Proposition 4.1. □

Proposition 4.4. *The collection $\{\mathcal{N}_b : b \in [0,1]\}$ is nonincreasing in b, and $\mathcal{N}_0 = \mathcal{N}$.*

Proof. These results are immediate from the definition of $\mathcal{N}_b$. □

Proposition 4.5. *For $\alpha \in [0,1)$ let $B_\alpha = \{b \in [0,1] : \mathcal{N}_b \cap C_\alpha \neq \emptyset\}$ and $b^*_\alpha = \sup B_\alpha$. Then $B_\alpha = [0, b^*_\alpha]$.*

Proof. Obviously $0 \in B_\alpha$ so B_α is nonempty, and hence b^*_α is well-defined. Let $b > 0$ with $b \in B_\alpha$. Let $b_1 \in [0,b)$. From Proposition 4.4, $\mathcal{N}_{b_1} \supseteq \mathcal{N}_b$, hence since $\mathcal{N}_b \cap C_\alpha \neq \emptyset$, then $\mathcal{N}_{b_1} \cap C_\alpha \neq \emptyset$. Therefore, $b_1 \in B_\alpha$. Let $\{b_n : n = 1, 2, \ldots\}$ be a sequence in B_α such that $b_n \uparrow b^*_\alpha$. For each $n = 1, 2, \ldots$ there exists an $\eta_n \in \mathcal{N}$ such that $\eta_n \in C_\alpha$ and $\sum_{m\in\mathcal{M}} \pi_m(\eta_{nm}) \geq Mb_n$. Consider the sequence $\{\eta_n\}$ in $\mathcal{N}$. This is a sequence belonging to the closed and bounded set C_α. By the Bolzano-Weierstrass Theorem (cf., (20)), there exists a subsequence $\{\eta_{n'}\}$ of $\{\eta_n\}$ such that for some $\eta_0 \in C_\alpha$, $\eta_{n'} \to \eta_0$. Furthermore, since the $\pi_m(\cdot)$s are continuous, then $\sum_{m\in\mathcal{M}} \pi_m(\eta_{0m}) = \lim_{n'\to\infty} \sum_{m\in\mathcal{M}} \pi_m(\eta_{n'm}) \geq M \lim_{n'\to\infty} b_{n'} = Mb^*_\alpha$. Therefore, $\eta_0 \in \mathcal{N}_{b^*_\alpha}$, hence $\eta_0 \in C_\alpha \cap \mathcal{N}_{b^*_\alpha}$. Thus, $b^*_\alpha \in B_\alpha$. □

Theorem 4.1. (Existence) *Let $\alpha \in [0,1)$. Then $C_\alpha \cap \mathcal{N}_{b^*_\alpha} \neq \emptyset$. Furthermore, $\eta \in \mathcal{N}$ is an FWER-α optimal size vector if and only if $\eta \in C_\alpha \cap \mathcal{N}_{b^*_\alpha}$.*

Proof. First, observe from Proposition 4.5 that $C_\alpha \cap \mathcal{N}_{b^*_\alpha} \neq \emptyset$. Each element $\eta_0 \in C_\alpha \cap \mathcal{N}_{b^*_\alpha}$ satisfies the FWER-α constraint and also achieves the optimal (largest) value of $\sum_{m\in\mathcal{M}} \pi_m(\eta_m)/M$ among all $\eta \in C_\alpha$. Therefore, η_0 is an optimal size vector for the M MP tests associated with the multiple hypothesis testing problem with FWER control at α.

Suppose that η_0 is an FWER-α optimal solution but $\eta_0 \notin C_\alpha \cap \mathcal{N}_{b^*_\alpha}$. Then there must exist a $b > b^*_\alpha$ such that $\eta_0 \in \mathcal{N}_b$. Since $\eta_0 \in C_\alpha$, then we have $C_\alpha \cap \mathcal{N}_b \neq \emptyset$. But this contradicts the maximality of b^*_α. Hence, the supposition could not be true. □

We remark that a crucial ingredient of the existence of an optimal size vector is that the constraint set C_α be a closed and convex set containing $\mathbf{0}$. This constraint set need not always be generated by the FWER constraint. For example, it could be a constraint which bounds the k-FWER such as $C_\alpha(k\text{-FWER}) = \{\eta \in \mathcal{N} : R_{0k}(\delta^*(\eta), Q_0) \le \alpha\}$, where $R_{0k}(\delta^*(\eta), Q_0) = \mathbf{P}\{\sum_{m\in\mathcal{M}} \delta^*_m(X_m, U_m; \eta_m) \ge k | X_m \sim Q_{m0}\}$. It is not difficult to see that an explicit form of this constraint set is

$$C_\alpha(k\text{-FWER}) = \left\{\eta \in \mathcal{N} : \sum_{t=0}^{k-1} \sum_{N\subset\mathcal{M};\ |N|=t} \left[\prod_{m\in N} \eta_m\right]\left[\prod_{m\in N^c}(1-\eta_m)\right] \ge 1-\alpha\right\}.$$

It is clear that this set contains $\mathbf{0}$ and that it is a closed set since the sum of continuous functions is also continuous. Therefore, if for a specified k, with $k \le M$, $C_\alpha(k\text{-FWER})$ is a convex set in $\mathcal{N}$, there will then exist an optimal size vector satisfying the constraint that the k-FWER is bounded above by α. This convexity question for $k > 1$ still needs resolution.

4.2. *Uniqueness of Optimal Size Vector.* Theorem 4.1 guarantees the existence of an optimal FWER size vector, but it does not address whether the solution is unique. We consider this uniqueness issue in this section. For this purpose, we first define the sections of C_α.

Definition 4.3. *Let $\alpha \in (0,1)$ and C_α be the constraint set. The mth section of C_α is the subset of $[0,1]$ given by $C_\alpha(m) = \{\eta_m \in [0,1] : \eta \in C_\alpha\}$.*

Theorem 4.2. (Uniqueness) *Let $\alpha \in [0,1)$. If, $\forall m \in \mathcal{M}, \eta_m \mapsto \pi_m(\eta_m)$ is strictly increasing on $C_\alpha(m)$, then the optimal FWER-α size vector is unique and it is η^* with $C_\alpha \cap \mathcal{N}_{b^*_\alpha} = \{\eta^*\}$.*

Proof. It suffices to show from Theorem 4.1 that $C_\alpha \cap \mathcal{N}_{b^*_\alpha}$ is a singleton set. Suppose it is not a singleton set. Let $\eta_1, \eta_2 \in \mathcal{N}$ with $\eta_1 \ne \eta_2$ such that for $j = 1,2$, $\sum_{m\in\mathcal{M}} \log(1-\eta_{jm}) \ge \log(1-\alpha)$ and $\sum_{m\in\mathcal{M}} \pi_m(\eta_{jm}) \ge M b^*_\alpha$. Let $\xi \in (0,1)$ and define $\eta^* = \xi\eta_1 + (1-\xi)\eta_2$. By convexity of both C_α and $\mathcal{N}_{b^*_\alpha}$, we have $\eta^* \in C_\alpha \cap \mathcal{N}_{b^*_\alpha}$. But, due to the strict concavity of the logarithmic map $\eta \mapsto \log(1-\eta)$, $\sum_{m\in\mathcal{M}} \log(1-\eta^*_m) = \sum_{m\in\mathcal{M}} \log[1-(\xi\eta_{1m} + (1-\xi)\eta_{2m})] > \xi \sum_{m\in\mathcal{M}} \log(1-\eta_{1m}) + (1-\xi)\sum_{m\in\mathcal{M}} \log(1-\eta_{2m}) \ge \log(1-\alpha)$. Thus, $\eta^* \in C_\alpha \setminus UB(C_\alpha)$. By continuity of the logarithm function, there exists an $m_0 \in \mathcal{M}$ and an $\epsilon_0 > 0$ such that $\sum_{m\in\mathcal{M}} \log[1-(\eta^*_m + \epsilon_0 I\{m = m_0\})] \ge \log(1-\alpha)$. Observe that $(\eta^*_m + \epsilon_0 I\{m = m_0\} : m \in \mathcal{M})$ belongs to both C_α and $U(\eta^*)$. Since $\forall m \in \mathcal{M}$, $\eta_m \mapsto \pi_m(\eta_m)$ is strictly increasing on

$C_\alpha(m)$, then $\sum_{m\in\mathcal{M}} \pi_m(\eta_m^* + \epsilon_0 I\{m = m_0\}) > \sum_{m\in\mathcal{M}} \pi_m(\eta_m^*) \geq M b_\alpha^*$. But this contradicts the maximality of b_α^*. Therefore, we must have $C_\alpha \cap \mathcal{N}_{b_\alpha^*}$ to be a singleton set, proving the uniqueness of the optimal solution. □

COROLLARY 4.1. *If, $\forall m \in \mathcal{M}$, $\eta_m \in [0, \sup C_\alpha(\eta_m)) \Rightarrow \pi_m(\eta_m) < 1$, then the optimal FWER-$\alpha$ size vector is unique.*

PROOF. This follows from Theorem 4.2 and Proposition 3.2 since the condition implies that, for $\forall m \in \mathcal{M}$, $\eta_m \mapsto \pi_m(\eta_m)$ is strictly increasing on $C_\alpha(m)$. □

Non-uniqueness of this optimal FWER size vector may occur when dealing with non-regular families of densities, such as the uniform or the shifted exponential densities, where the power may equal one even though the size is still less than one. Thus, the mapping $\eta \mapsto \pi(\eta)$ in such situations is not strictly increasing, and there need not be a unique optimal size vector. In such cases, the set $C_\alpha \cap \mathcal{N}_{b_\alpha^*}$ is not a singleton set.

5. Finding the Optimal Sizes. In this section we consider the computational problem of finding the optimal FWER size vector. Generally, without differentiability of the power functions (as functions of their sizes) such as in problems dealing with discrete distributions, there may be a need to invoke linear or nonlinear programming methods to obtain the optimal solution. Below we present an analytic method for obtaining the optimal sizes under the condition that for each $m \in \mathcal{M}$, $\pi_m(\eta_m)$ is twice-differentiable with respect to η_m.

THEOREM 5.1. *Suppose that, for each $m \in \mathcal{M}$, the power $\pi_m(\eta_m)$ of the size η_m Neyman-Pearson most powerful test $\delta_m^*(\eta_m)$ of H_{m0} versus H_{m1} is twice-differentiable with respect to η_m with continuous positive first derivative $\pi'_m(\eta_m)$ and second derivative $\pi''_m(\eta_m)$. Let $\eta = (\eta_1, \eta_2, \ldots, \eta_M) \in \mathcal{N}$ be a solution to the set of (Lagrange) equations*

$$\forall m \in \mathcal{M}, \ \pi'_m(\eta_m)(1 - \eta_m) = \lambda, \ \textit{for some } \lambda \in \Re_+; \tag{5.1}$$

$$\textstyle\sum_{m\in\mathcal{M}} \log(1 - \eta_m) = \log(1 - \alpha). \tag{5.2}$$

Then, η is the optimal size vector for controlling the FWER at level α.

PROOF. First, observe that since $\pi'_m(\cdot)$s are all positive, then by Theorem 4.2 there is a unique optimal solution. Now, form the Lagrange function

$$J(\eta_1, \eta_2, \ldots, \eta_M, \lambda) = \sum_{m\in\mathcal{M}} \pi_m(\eta_m) + \lambda \left\{ \sum_{m\in\mathcal{M}} \log(1 - \eta_m) - \log(1 - \alpha) \right\}.$$

Then,

$$\frac{\partial J}{\partial \eta_m} = \pi'_m(\eta_m) - \frac{\lambda}{1-\eta_m}, \; m \in \mathcal{M};$$
$$\frac{\partial J}{\partial \lambda} = \sum_{m=1}^{M} \log(1-\eta_m) - \log(1-\alpha).$$

Equating these to zeros, we obtain conditions (5.1) and (5.2) in the statement of the theorem. Thus, a solution $(\eta_1, \ldots, \eta_M, \lambda)$ of (5.1) and (5.2) is an extremum point of the Lagrange function.

To show that the solution of (5.1) and (5.2) is the maximizer of the Lagrange function J, we need to verify that the sequence of determinants of the principal minors of the bordered Hessian matrix, evaluated at this solution, alternates in signs. The second partial derivatives of the Lagrange function are

$$\frac{\partial^2 J}{\partial \eta_m \partial \eta_n} = \left\{\pi''_m(\eta_m) - \frac{\lambda}{(1-\eta_m)^2}\right\} I\{m=n\}, \; m,n \in \mathcal{M};$$
$$\frac{\partial^2 J}{\partial \eta_m \partial \lambda} = \frac{\partial^2 J}{\partial \lambda \partial \eta_m} = -\frac{1}{1-\eta_m}, \; m \in \mathcal{M};$$
$$\frac{\partial^2 J}{\partial \lambda^2} = 0.$$

The solution $(\eta_1, \eta_2, \ldots, \eta_M, \lambda)$ of the Lagrange equations (5.1) and (5.2) satisfies $\lambda = \pi'_m(\eta_m)(1-\eta_m), \; m \in \mathcal{M}$. Since $\pi'_m(\eta_m) > 0$, then, at the solution, $\lambda \geq 0$. Furthermore, since as shown in Proposition 3.1, $\pi_m(\cdot)$ is concave, then $\pi''_m(\eta_m) \leq 0$. As a consequence, at the solution,

$$\frac{\partial^2 J}{\partial \eta_m \partial \eta_n} \leq 0, \quad \frac{\partial^2 J}{\partial \eta_m \partial \lambda} < 0, \quad \frac{\partial^2 J}{\partial \lambda^2} = 0.$$

The associated bordered $(M+1) \times (M+1)$ Hessian matrix evaluated at the solution of (5.1) and (5.2) is of form

$$\mathbf{H} = -\begin{bmatrix} \mathrm{Dg}(\mathbf{b}) & \mathbf{a} \\ \mathbf{a}^t & 0 \end{bmatrix}, \tag{5.3}$$

where $\mathrm{Dg}(\mathbf{b})$ is the diagonal matrix with diagonal elements consisting of the elements of the vector $\mathbf{b}$ and with $\mathbf{b}^t = -\left(\pi''_m(\eta_m) - \lambda/(1-\eta_m)^2, \; m \in \mathcal{M}\right)$ and $\mathbf{a}^t = \left(1/(1-\eta_m), \; m \in \mathcal{M}\right)$. Observe that all the elements of $\mathbf{b}$ and $\mathbf{a}$ are nonnegative. The mth principal minor of $\mathbf{H}$ is

$$\mathbf{H}_m = \begin{bmatrix} \mathrm{Dg}(\mathbf{b}_m) & \mathbf{a}_m \\ {\mathbf{a}_m}^t & 0 \end{bmatrix} \tag{5.4}$$

where $\mathbf{b}_m = (b_1, b_2, \ldots, b_m)^t$ and $\mathbf{a}_m = (a_1, a_2, \ldots, a_m)^t$. Using the general result in Lemma 5.1 presented below, we find that

$$\det(\mathbf{H}_m) = (-1)^{m+2} \left(\prod_{k=1}^{m} b_k \right) \sum_{k=1}^{m} \frac{a_k^2}{b_k}. \tag{5.5}$$

Since the a_ks and b_ks are nonnegative, it follows that the determinants of the principal minors of the bordered Hessian matrix alternate in sign, starting with a negative sign. Consequently, the solution of (5.1) and (5.2) is a maximizer of the Lagrange function, and hence maximizes $\frac{1}{M} \sum_{m \in \mathcal{M}} \pi_m(\eta_m)$ subject to the FWER-α level constraint. □

We now establish the determinantal lemma used in the preceding proof.

LEMMA 5.1. *For any $M \times 1$ vectors* $\mathbf{b}$ *and* $\mathbf{a}$,

$$\det \begin{bmatrix} Dg(\mathbf{b}) & \mathbf{a} \\ \mathbf{a}^t & 0 \end{bmatrix} = -\left(\prod_{m \in \mathcal{M}} b_m \right) \sum_{m \in \mathcal{M}} \frac{a_m^2}{b_m}.$$

PROOF. The claimed result clearly holds for $M = 1$. Assume that it holds for $M = 1, 2, \ldots, K$. We show that it holds for $M = K + 1$. Let $\mathbf{a} = (a_1, a_2, \ldots, a_K, a_{K+1})^t = (a_1, \mathbf{a}^*)$ and $\mathbf{b} = (b_1, b_2, \ldots, b_K, b_{K+1})^t = (b_1, \mathbf{b}^*)$. Then

$$\begin{aligned}
&\det \begin{bmatrix} b_1 & \mathbf{0}^t & a_1 \\ \mathbf{0} & \mathrm{Dg}(\mathbf{b}^*) & \mathbf{a}^* \\ a_1 & \mathbf{a}^{*t} & 0 \end{bmatrix} \\
&= (-1)^2 b_1 \det \begin{bmatrix} \mathrm{Dg}(\mathbf{b}^*) & \mathbf{a}^* \\ \mathbf{a}^{*t} & 0 \end{bmatrix} + (-1)^{K+3} a_1 \det \begin{bmatrix} \mathbf{0} & \mathrm{Dg}(\mathbf{b}^*) \\ a_1 & \mathbf{a}^{*t} \end{bmatrix} \\
&= b_1 (-1) \left(\prod_{m=2}^{K+1} b_m \right) \left(\sum_{m=2}^{K+1} \frac{a_m^2}{b_m} \right) + a_1 (-1)^{K+3} (-1)^K \det \begin{bmatrix} \mathrm{Dg}(\mathbf{b}^*) & \mathbf{0} \\ \mathbf{a}^{*t} & a_1 \end{bmatrix} \\
&= -\left(\prod_{m=1}^{K+1} b_m \right) \left(\sum_{m=2}^{K+1} \frac{a_m^2}{b_m} \right) - \left(\prod_{m=1}^{K+1} b_m \right) \left(\frac{a_1^2}{b_1} \right) \\
&= = -\left(\prod_{m=1}^{K+1} b_m \right) \left(\sum_{m=1}^{K+1} \frac{a_m^2}{b_m} \right).
\end{aligned}$$

By the principle of mathematical induction, this proves the lemma. □

We establish a monotonocity property of the optimal FWER-controlling size vector. We prove the result under the differentiability conditions in Theorem 5.1.

Proposition 5.1. *Assume the conditions of Theorem 5.1. Then, for each $m \in \mathcal{M}$, the mapping $\alpha \mapsto \eta_m(\alpha)$ is nondecreasing.*

Proof. For each $m \in \mathcal{M}$, define the mapping from $[0,1]$ to $\Re$ via

$$\eta_m \mapsto g_m(\eta_m) \equiv \pi_m'(\eta_m)(1-\eta_m).$$

Since $\pi_m'(\eta_m) \geq 0$ by nondecreasing property, then $g_m(\eta_m) \geq 0, \forall \eta_m \in [0,1]$. Furthermore,

$$g_m'(\eta_m) = \pi_m''(\eta_m)(1-\eta_m) - \pi_m'(\eta_m),$$

so since $\pi_m''(\eta_m) \leq 0$ by concavity, then $g_m'(\eta_m) \leq 0$. Therefore, each $g_m(\cdot)$ is a nonincreasing function. Now, the defining condition of the optimal FWER-controlling procedure in (5.1) can be re-stated in terms of these g_ms via $g_m(\eta_m) = \lambda, \ \forall m \in \mathcal{M}$, together with the constraint condition in (5.2). Suppose that we increase the value of α from α_1 to α_2. This will entail that the left-hand side of (5.2) must decrease, but in order for this to happen, and since each $g_m(\cdot)$ is nonincreasing, we must decrease the common value of the Lagrange constant λ. But, by doing so, each of the η_m cannot decrease. This shows that for each $m \in \mathcal{M}$, we must have $\eta_m(\alpha_1) \leq \eta_m(\alpha_2)$, thereby establishing the nondecreasing property of each of the $g_m(\cdot)$s. □

The result in Proposition 5.1 is actually a desirable property since this implies that if at FWER size α_1, $\delta_m(\eta_m(\alpha_1))$ resulted in the rejection of H_{m0}, then at an $\alpha_2 > \alpha_1$, $\delta_m(\eta_m(\alpha_2))$ will certainly also lead to rejection of H_{m0}. This result is also critical in proving a martingale property needed for the extension to an FDR-controlling procedure.

6. When Densities Possess MLR Property. At this point we discuss a perceived potential limitation of the optimization framework considered above which deals with the case of a simple null hypothesis versus a simple alternative hypothesis for each $m \in \mathcal{M}$. This may seem too restrictive a framework since in most cases one will be dealing with composite hypotheses. Consider then the situation where, for each $m \in \mathcal{M}$, the density function q_m belongs to a one-dimensional family $\mathcal{F}_m = \{q_m(\cdot;\theta_m) : \ \theta_m \in \Theta_m \subset \Re\}$ with this family possessing the monotone likelihood ratio (MLR) property, cf., (16). A typical pair of hypotheses to be tested in this situation will be $H_{m0}^* : \theta_m \leq \theta_{m0}$ versus $H_{m1}^* : \theta_m > \theta_{m0}$. Because of the MLR property, then there is a uniformly most powerful (UMP) test $\delta_m(X_m;\eta_m)$ of size η_m for this problem, and this UMP test is exactly of form of the MP test of size η_m for the simple null hypothesis $H_{m0} : \theta_m = \theta_{m0}$ versus the simple alternative hypothesis $H_{m1} : \theta_m = \theta_{m1}$, with $\theta_{m1} > \theta_{m0}$.

Now, in practice, the usual approach will be for the investigator to specify the value of θ_{m1} at which value the power is of interest, or a value which is deemed scientifically, e.g., clinically, relevant. The specification of such a θ_{m1} is usually equivalent to specifying an *effect size*. This is the approach utilized, for instance, in sample size determination problems. We may then denote by $\pi_m(\eta_m)$ the power at $\theta_m = \theta_{m1}$ of the UMP test of size η_m, and in the multiple hypotheses testing problem, the goal will be to maximize $\frac{1}{M}\sum_{m\in\mathcal{M}}\pi_m(\eta_m)$ subject to the constraint that the FWER, which equals $1-\prod_{m\in\mathcal{M}}(1-\eta_m)$, does not exceed a pre-specified level α. The problem then reverts back to that considered with the simple null versus simple alternative hypotheses. Thus, the framework is not after all too limiting.

7. In Relation to p-Value Statistics. Let us also relate the results to the p-value or significance value distributions. From Corollary 3.2 the power function $\pi_m(\eta_m)$ is differentiable if and only if the distribution function of $S_m(X_m)$ under $H_{m1}: Q_m = Q_{m1}$ is differentiable, and in this case the derivative $\pi_m'(\cdot)$ of the power function $\pi_m(\cdot)$ coincides with the density function of the p-value statistic $S_m(X_m)$ under $H_{m1}: Q_m = Q_{m1}$. Therefore, denoting by $h_m(\cdot)$ this density function, $\pi_m'(\eta_m) = h_m(\eta_m)$, so the first condition (5.1) in Theorem 5.1 may be restated in terms of the density functions, under H_{m1}s, of the p-value statistics $S_m(X_m)$s. This first condition on the optimal FWER size vector becomes

$$h_m(\eta_m)(1-\eta_m) = \text{Constant}, \ \forall m \in \mathcal{M}. \tag{7.1}$$

This is a surprising result since this indicates that it is not enough to simply find the sizes that maximizes these p-value statistics density functions under the alternative hypotheses, as is dictated by the Neyman-Pearson Lemma when simply dealing with a single pair of null and alternative hypotheses. Rather, in the multiple hypotheses testing setting, some attenuation is required in the sense that a large size incurs some penalty. This phenomenon will be later referred to as a *size-investing strategy* in this multiple testing setting.

We also point out that the resulting optimal FWER-controlling MHTDF δ^* may still be converted to a procedure based on the p-value statistics, though in the general case where the X_ms may be discrete, there is a need to utilize the notion of randomized p-value statistics. In the case of continuous X_ms so the Neyman-Pearson MP tests do not require randomization, if $\eta = (\eta_1, \ldots, \eta_M)$ is the optimal FWER-α size vector and $(p_1(x_1), \ldots, p_M(x_M))$ is the vector of computed p-values, the decision or action based on the data

$\mathbf{x} = (x_1, \ldots, x_M) \in \mathcal{X} = \mathcal{X}_1 \times \ldots \times \mathcal{X}_M$ is

$$\delta^*(\mathbf{x}) = (I\{p_1(x_1) \leq \eta_1\}, I\{p_2(x_2) \leq \eta_2\}, \ldots, I\{p_M(x_M) \leq \eta_M\}).$$

Observe that in this setting, it is possible that $p_{m_1}(x_{m_1}) < p_{m_2}(x_{m_2})$ but $H_{m_1 0}$ is not rejected whereas $H_{m_2 0}$ is rejected depending, on the values of η_{m_1} and η_{m_2}. In a sense the decision-making with regards to the M pairs of hypotheses is not transitive with respect to the p-value statistics.

8. Concrete Examples for FWER Control. We demonstrate the FWER-control optimality results obtained in the preceding sections by considering three concrete situations. The first example deals with normal distributions, the second example uses exponential distributions. Both of these concrete settings allow the Lagrange approach. The third example uses binomial distributions and due to the non-differentiability of the power function, as a function of the size, does not admit the Lagrange approach.

8.1. *Testing for Normal Means.* For $m \in \mathcal{M}$ let $X_m \sim N(\mu_m, \sigma^2_{m0})$ where the μ_ms are unknown and the σ^2_{m0}s are known. Consider the multiple hypotheses testing problems where, based on X_m, we would like to test $H_{m0} : \mu_m = \mu_{m0}$ versus $H_{m1} : \mu_m = \mu_{m1}$ with $\mu_{m0} < \mu_{m1}$ for each $m \in \mathcal{M}$. Then the Neyman-Pearson MP test of size η_m for H_{m0} versus H_{m1} is

$$\delta_m^{NP}(X_m; \eta_m) \equiv \delta_m^{NP}(\eta_m) = I\{X_m \geq \mu_{m0} + \sigma_{m0}\Phi^{-1}(1 - \eta_m)\}, \tag{8.1}$$

where $\Phi(\cdot)$ and $\Phi^{-1}(\cdot)$ are the cumulative distribution and quantile functions of a standard normal random variable. Defining the mth effect size by $\gamma_m = (\mu_{m1} - \mu_{m0})/\sigma_{m0}$, then the power function of $\delta_m^{NP}(\eta_m)$ is given by

$$\pi_m^{NP}(\eta_m) = \Phi(\gamma_m - \Phi^{-1}(1 - \eta_m)). \tag{8.2}$$

Consequently, with $\phi(z) = \exp(-z^2)/\sqrt{2\pi} = \Phi'(z)$ the standard normal density function, we obtain the derivative of this power function with respect to η_m to be

$$(\pi_m^{NP})'(\eta_m) = \frac{\phi(\gamma_m - \Phi^{-1}(1 - \eta_m))}{\phi(\Phi^{-1}(1 - \eta_m))}. \tag{8.3}$$

For a fixed $\alpha \in (0, 1)$ and γ_ms, consider the mappings $d \mapsto \eta_m(d), m \in \mathcal{M}$, defined implicitly via the equation

$$\frac{\phi(\gamma_m - \Phi^{-1}(1 - \eta_m))}{\phi(\Phi^{-1}(1 - \eta_m))}(1 - \eta_m) - d = 0. \tag{8.4}$$

TABLE 1

Optimal test sizes under normality for different power/effect sizes configurations. The configuration and the optimal sizes are described by the notation $k:(a,b,...)$ which is interpreted as having k of each of the elements in the vector $(a,b,...)$. The relative efficiency (in percent) of the optimal procedure relative to the Sidak procedure is also presented.

Effect Size, γ, Configuration	Optimal Test Sizes/[Efficiency over Sidak (in %)]	
	$M=4$	$M=20$
$M:1$	$4:.0127$ [100.0]	$20:.0026$ [100.0]
$M/2:(.5,1)$	$2:(.0009,.0245)$ [113.6]	$10:(0,.0051)$ [125.1]
$M/2:(1,2)$	$2:(.0050,.0245)$ [104.5]	$10:(.0001,.0050)$ [115.3]
$M/2:(1,5)$	$2:(.0228,.0026)$ [103.6]	$10:(.0035,.0016)$ [100.3]
$M/4:(0.5,1,2,4)$	$1:(.0001,.0128,.0303,.0075)$ [105.4]	$5:(0,.0003,.0068,.0031)$ [107.1]
$M/4:(1,2,4,8)$	$1:(.0128,.0304,.0075,0)$ [105.0]	$5:(.0003,.0068,.0031,0)$ [104.3]

The optimal value of d, denoted by d^*, solves the equation

$$\sum_{m=1}^{M}\log(1-\eta_m(d))-\log(1-\alpha)=0. \tag{8.5}$$

The optimal sizes of the M Neyman-Pearson MP tests are the values of $\eta_m(d^*), m\in\mathcal{M}$. An `R` ((13)) implementation of this numerical problem first defines $v_m=1-\Phi^{-1}(1-\eta_m)$, so that condition (8.4) amounts to solving for $v_m=v_m(d)$ the equation

$$\log\Phi(v_m)+\gamma_m v_m-\log(d)-\gamma_m^2/2=0. \tag{8.6}$$

The `R` implementation utilizes a Newton-Raphson iteration in solving for v_ms in (8.6) and the `uniroot` routine in the `R` Library to solve for d in (8.5). After obtaining the $v_m(d)$s, the $\eta_m(d)$s are computed via $\eta_m(d)=1-\Phi(v_m(d))$.

Figures 1 and 2 demonstrate the optimal sizes for different effect sizes when $M=2000$ for uniformly distributed effect sizes and folded-normal distributed effect sizes. Observe that when the effect size is small, which converts to low power, then the optimal size for the test is also small, but also note that when the effect size is large, which converts to high power, then the optimal test size is also small. For the tests with moderate effect sizes or power, then the optimal sizes are higher. This behavior could also be seen by looking at the second panels in each figure which show the achieved power of the tests at the optimal sizes. Observe that when the powers pass 0.50, then the test sizes start decreasing. This behavior could also be observed by examining Table 1 for small values of M.

We also compared the efficiency of the optimal procedure relative to the Sidak procedure. This measure of efficiency is the ratio (multiplied by 100) of the average power over the M tests, defined by $\sum_{m \in \mathcal{M}} \pi_m(\eta_m)/M$, of the optimal procedure and the average power for the Sidak procedure. The fourth panels in Figures 1 and 2 depict the powers of the resulting tests versus the effect size for both procedures (solid blue = optimal; dashed red=Sidak). In the uniformly-generated effect sizes, the efficiency of the optimal procedure over the Sidak is 103.5%. As can be seen in Table 1, this efficiency is affected by the vector of effect sizes. For instance, when we change the effect sizes in Figure 1 to be generated from a uniform over $[.1, 2]$, then the efficiency jumps to 181.7%, though it should also be pointed out that since the effect sizes are small, then the overall powers of both procedures are also small. On the otherhand, for the folded-normal-generated effect sizes, the efficiency is 110.9%.

8.2. *Testing for Exponential Rates.* In this subsection we focus our attention to the situation where the responses are distributed according to an exponential distribution. For $m \in \mathcal{M}$, let $X_{mi}, i = 1, 2, \ldots, n$, be IID from an exponential distribution with mean $1/\lambda_m$ or rate λ_m. We consider multiple testing problem with hypotheses $H_{m0} : \lambda_m = \lambda_{m0}$ versus $H_{m1} : \lambda_m = \lambda_{m1}$. It is assumed that $\lambda_{m0} < \lambda_{m1}$. Denote by $S_m = \sum_{i=1}^n X_{mi}$, which is the sufficient statistic for λ_m. The Neyman-Pearson size η_m MP test for H_{m0} versus H_{m1} is

$$\delta_m^{NP}(S_m; \eta_m) = I\{2\lambda_{m0} S_m \leq c_m(\eta_m)\} \tag{8.7}$$

where $c_m(\eta_m) = G_{2n}^{-1}(\eta_m)$ with $G_k(\cdot)$ and $G_k^{-1}(\cdot)$ denoting, respectively, the cumulative distribution and quantile functions for a chi-squared random variable with k degrees-of-freedom. The appropriate effect size for the mth testing problem is $\rho_m = \lambda_{m1}/\lambda_{m0}$. The power function of δ_m^{NP} as a function of the size η_m is $\pi_m^{NP}(\eta_m) = G_{2n}\left(\rho_m G_{2n}^{-1}(\eta_m)\right)$, and the derivative of this function with respect to η_m is

$$(\pi_m^{NP})'(\eta_m) = \rho_m \frac{g_{2n}(\rho_m G_{2n}^{-1}(\eta_m))}{g_{2n}(G_{2n}^{-1}(\eta_m))} \tag{8.8}$$

where $g_k(\cdot)$ is the density function associated with $G_k(\cdot)$, which is given by

$$g_k(w) = \frac{1}{2^{k/2}\Gamma(k/2)} w^{k/2-1} \exp(-w/2) I\{w \geq 0\},$$

Fig 1. *Optimal test sizes and powers for 2000 MP tests of hypotheses under normality when the effect sizes were generated from a uniform*$[.1, 10]$ *distribution. Panel four shows the powers for both the optimal [solid black] and the Sidak [dashed red] tests with respect to effect sizes.*

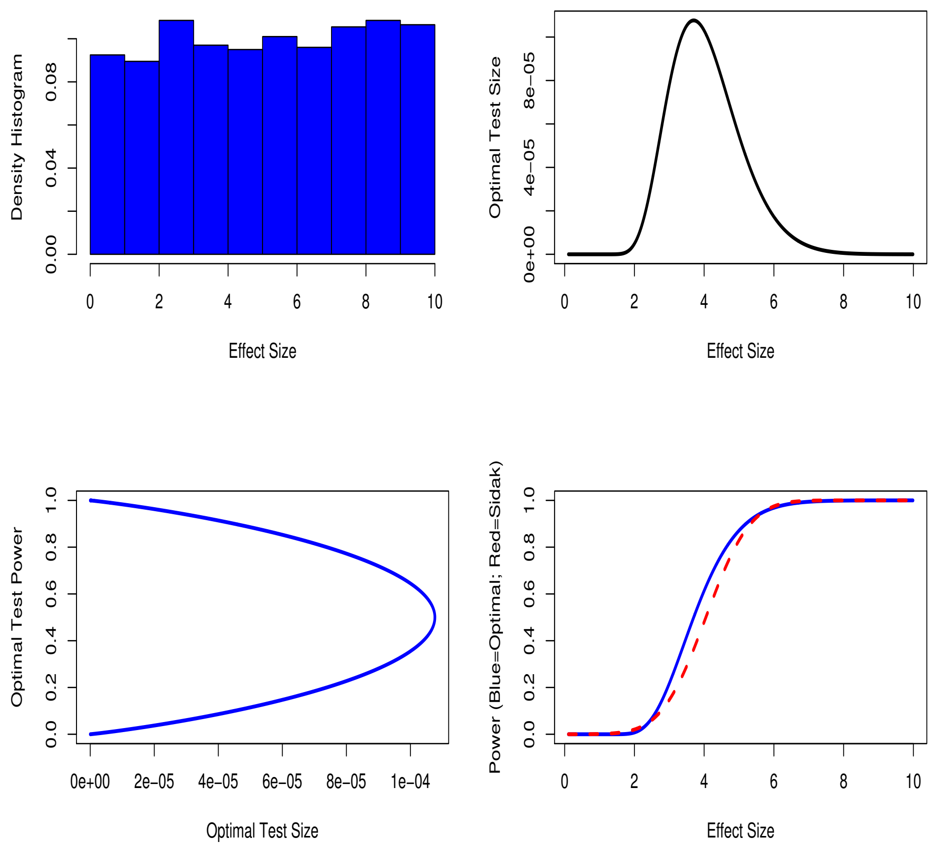
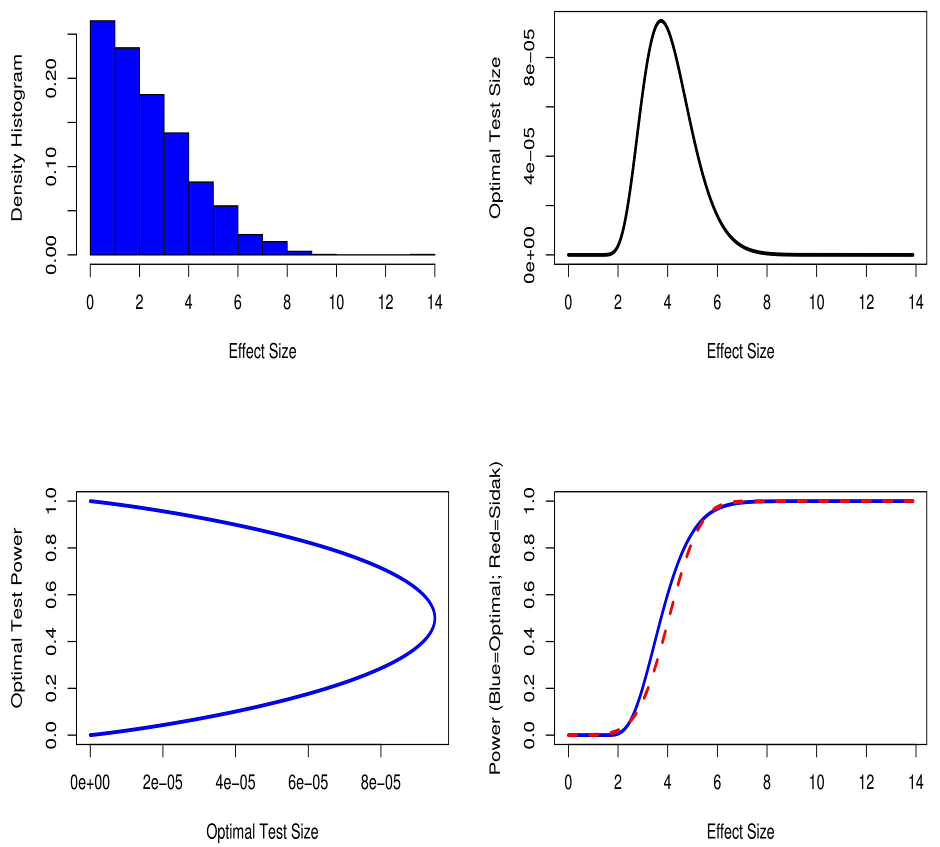


Fig 2. *Optimal test sizes and powers for 2000 MP tests of hypotheses under normality when the effect sizes were generated from a folded normal distribution*$(0, 3)$ *distribution.*



with $\Gamma(\cdot)$ the gamma function. The equations for determining the optimal test sizes are

$$\rho_m(1-\eta_m)\rho_m g_{2n}(\rho_m G_{2n}^{-1}(\eta_m))/g_{2n}(G_{2n}^{-1}(\eta_m)) = d, \quad m \in \mathcal{M};$$
$$\textstyle\sum_{m\in\mathcal{M}} \log(1-\eta_m) - \log(1-\alpha) = 0.$$

As in the normal testing problem, a variable substitution $v_m = G_{2n}^{-1}(\eta_m), m \in \mathcal{M}$, facilitates the computation of the η_ms. Noting that $g_{2n}(\rho v)/g_{2n}(v) = \rho^{n-1}\exp\{-v(\rho-1)/2\}$, the problem is to solve in d the equations

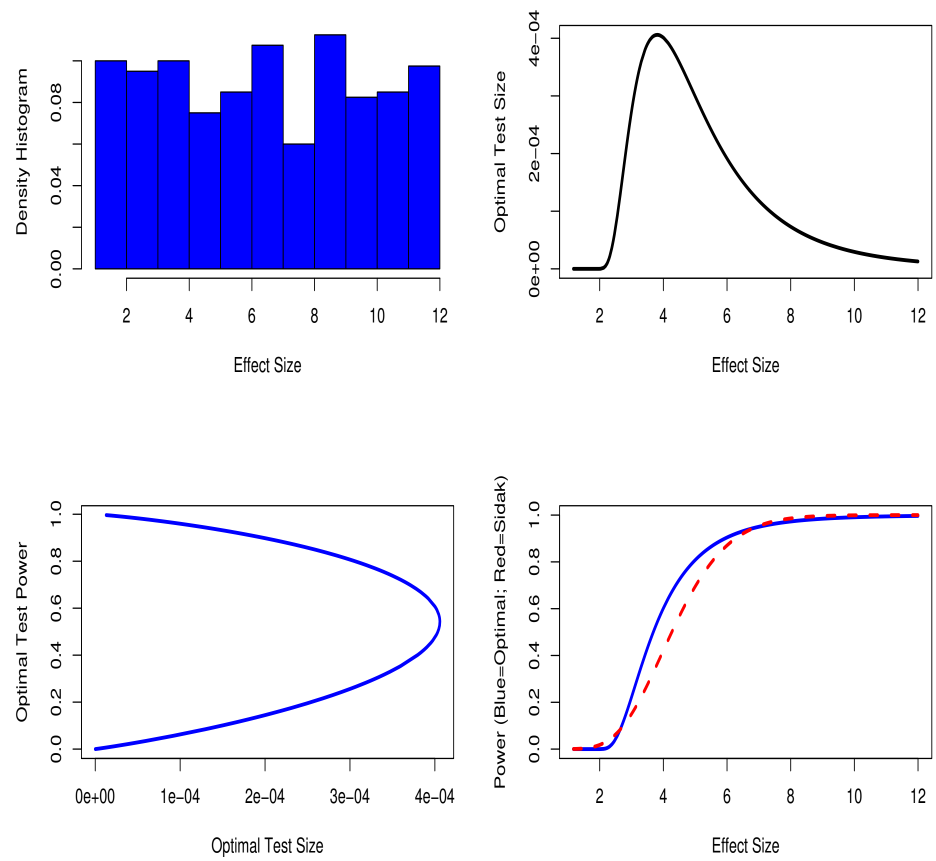
$$n\log(\rho_m) + log[1-G_{2n}(v_m)] - (\rho_m - 1)v_m/2 - \log(d) = 0, \quad m \in \mathcal{M}; \tag{8.9}$$
$$\sum_{m\in\mathcal{M}} \log[1-G_{2n}(v_m)] - \log(1-\alpha) = 0. \tag{8.10}$$

Having obtained the value of d, say d^*, that satisfies these equations, the optimal test sizes are computed via $\eta_m^* = G_{2n}(v_m(d^*))$. The value of v_ms given d in (8.9) are again obtained using a Newton-Raphson iteration as in the normal testing setting. As in the normal testing problem, this procedure for the exponential setting was implemented via an `R` program. Figure 3 shows the optimal test size values in the second plot frame when $M = 400$ and when the effect sizes ρ_ms were generated from a uniform distribution over $[1.1, 12]$. The sample size utilized was $n = 10$. The first plot frame is the density histogram of the effect sizes, while the third plot frame shows the powers of the tests at the optimal test sizes. We observe here the same pattern that occurred in the normal test setting, which is that when the test power is either very low or very high, then the optimal test size becomes very low as well. However, in contrast to the normal testing problem, the test power at which the optimal size changes from increasing to decreasing is not equal to 0.50.

We also computed the efficiency of the optimal procedure relative to the Sidak procedure, and in this case it was 104.1%. Again, this efficiency depends to a large extent on the effect size vector. In this exponential example, when we generate the effect sizes to be from a uniform over $[1.1, 2]$ with everything else the same as in Figure 3, the efficiency jumps to 186.9%, though we should point out that because the effect sizes are small, then the resulting overall powers for both the optimal and Sidak procedures are small.

8.3. *Testing for Bernoulli Parameters.* We now consider a discrete distribution which leads to the situation where the power functions are non-differentiable with respect to the sizes. This renders Theorem 5.1 not applicable. We assume that for each $m \in \mathcal{M}$, $X_m \sim BIN(n_m, \theta_m)$, with

FIG 3. *Optimal test sizes and powers for 400 MP tests of hypotheses under exponentiality when the effect sizes were generated from a uniform distribution on the interval 1.1 to 12. The sample size was* $n = 10$.



the n_m known. We consider the multiple hypotheses problem of testing $H_{m0} : \theta_m = \theta_{m0}$ versus $H_{m1} : \theta_m = \theta_{m1}$ for $m \in \mathcal{M}$, with $\theta_{m1} > \theta_{m0}$. The MP test of size η_m based on X_m for testing H_{m0} versus H_{m1} is

$$\delta_m^{NP}(\eta_m) = \begin{cases} 1 & \text{if} \quad X_m > c_m(\eta_m) \\ \gamma_m(\eta_m) & \text{if} \quad X_m = c_m(\eta_m) \\ 0 & \text{if} \quad X_m < c_m(\eta_m) \end{cases},$$

where, with $B(\cdot; n, \theta)$, $b(\cdot; n, \theta)$, and $B^{-1}(\cdot; n, \theta)$ denoting, respectively, the cumulative, probability mass, and quantile functions of a binomial distribution with parameters n and θ,

$$\begin{aligned} c_m(\eta_m) &= B^{-1}(1 - \eta_m; n_m, \theta_{m0}); \\ \gamma_m(\eta_m) &= \frac{B(c_m(\eta_m); n_m, \theta_{m0}) - (1 - \eta_m)}{b(c_m(\eta_m); n_m, \theta_{m0})}. \end{aligned}$$

The power functions are, for $m \in \mathcal{M}$,

$$\pi_m(\eta_m) = 1 - B(c_m(\eta_m); n_m, \theta_{m1}) + \gamma_m(\eta_m) b(c_m(\eta_m); n_m, \theta_{m1}). \tag{8.11}$$

Observe that because $B(\cdot; n_m, \theta_{m0})$ is a right-continuous nondecreasing step function, then $\eta_m \mapsto c_m(\eta_m)$ is a left-continuous nonincreasing step-function. However, as was established in Proposition 3.1, $\eta_m \mapsto \pi_m(\eta_m)$ is a nondecreasing continuous function, aside from being concave. Thus, since $\eta_m \mapsto \gamma_m(\eta_m)$ is a piecewise linear function, then the power function is piecewise linear. Therefore, it is piecewise differentiable, but at the cusps, its left- and right-hand derivatives are not equal, hence it is not differentiable at all points. These properties can be seen in Figure 4 which is a plot of the power function when $n_1 = 5$, $\theta_{10} = 0.50$, and $\theta_{11} = 0.70$. As a consequence of these properties, for each $b \in [0, 1]$, the subset of $\mathcal{N} = [0, 1]^M$ defined via $\mathcal{N}_b = \{\eta \in \mathcal{N} : \sum_{m \in \mathcal{M}} \pi_m(\eta_m) \geq Mb\}$ is a convex polyhedron. According to the theory developed earlier, the computational problem is to find the largest $b = b_\alpha^*$ such that the intersection between $\mathcal{N}_{b_\alpha^*}$ and the constraint set upper boundary $UB(C_\alpha) = \{\eta \in \mathcal{N} : \sum_{m \in \mathcal{M}} \log(1 - \eta_m) = \log(1 - \alpha)\}$ is a singleton point, with this point of intersection being the vector of optimal sizes. The computational difficulty in these discrete-type settings is that this point of intersection may occur in an edge of the convex polyhedron $\mathcal{N}_b$, where partial derivatives do not exist, instead of in a face of this polyhedron. The Lagrange approach to the solution will not work in this situation. For our illustration, since we know that the solution belongs to $UB(C_\alpha)$, we first express η_M in terms of $\eta_1, \ldots, \eta_{M-1}$, which will be

$$\eta_M(\eta_1, \ldots, \eta_{M-1}) = 1 - \frac{1 - \alpha}{\prod_{m \neq M}(1 - \eta_m)},$$

FIG 4. *Power function of the most powerful test for a Bernoulli parameter as a function of its size* η *when* $n = 5$, $\theta_0 = 0.5$, *and* $\theta_1 = 0.7$.

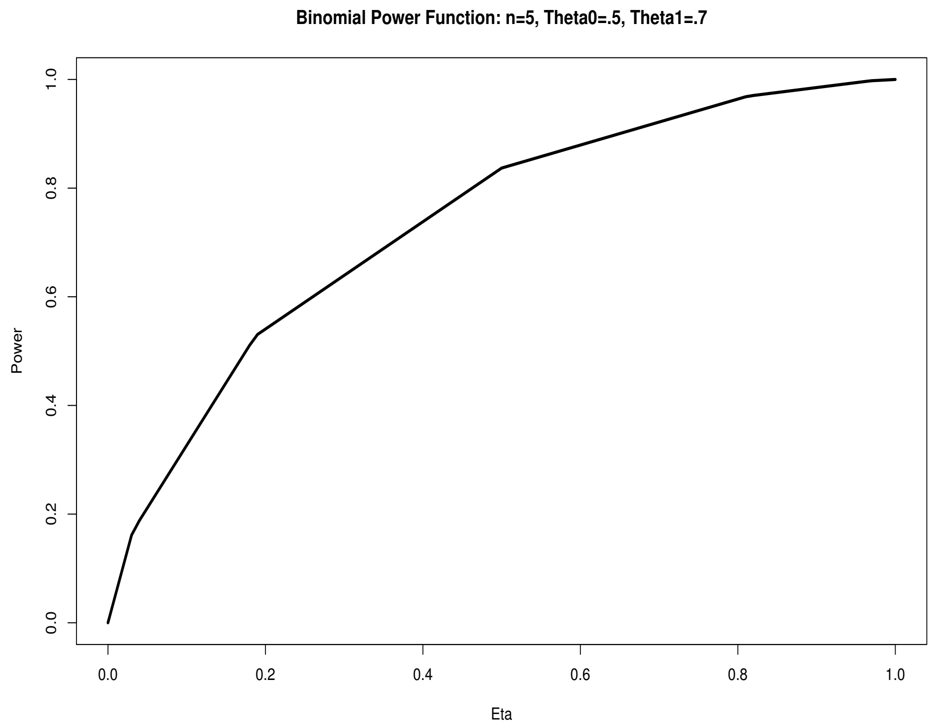


TABLE 2
Illustrative example for the multiple testing problem for Bernoulli/binomial distributions with $M = 5$ and $n_m = 5$ for $m = 1, 2, 3, 4, 5$ and $\alpha = .05$. The column 'Optimal Sizes' and 'Optimal Powers' refer to the sizes and powers of the tests. Also included are the sizes and powers of the Sidak procedure.

m	θ_{m0}	θ_{m1}	Optimal Sizes	Optimal Powers	Sidak Sizes	Sidak Powers
1	0.4	0.60	0.0102	0.0774	0.0102	0.0775
2	0.5	0.60	0.0000	0.0002	0.0102	0.0253
3	0.3	0.70	0.0307	0.5282	0.0102	0.2668
4	0.3	0.50	0.0043	0.0416	0.0102	0.0741
5	0.4	0.55	0.0053	0.0262	0.0102	0.0501
Total	NA	NA	NA	0.6738	NA	0.4940

and then use the `optim` function in `R` to optimize the objective function given by

$$o(\eta_1, \ldots, \eta_{M-1}) = \frac{1}{M}\left[\sum_{m=1}^{M-1} \pi_m(\eta_m) + \pi_M(\eta_M(\eta_1, \ldots, \eta_{M-1}))\right],$$

subject to the constraint that $\eta_m \in [0, 1]$ for $m \in \mathcal{M}$. To illustrate, we implemented this computational procedure for $M = 5$, $n_m = 5$ for $m = 1, 2, \ldots, 5$, and with null hypotheses values $\theta_0 = (.4, .5, .3, .3, .4)$ and alternative hypotheses values $\theta_1 = (.6, .6, .7, .5, .55)$. The FWER threshold was set to $\alpha = .05$. Table 2 provides the resulting optimal sizes, together with the power of each of the MP tests at these optimal sizes.

Note that these set of sizes for both the optimal and the Sidak procedures lead to an FWER equal to 0.05. As in the normal and exponential settings, we obtain the efficiency of the optimal procedure relative to the Sidak procedure via $(0.6738/0.4940) \times 100 = 140.4\%$.

We considered another set of parameters in this Bernoulli setting where we took $M = 10$, $n_m = 5$ for each $m = 1, 2, \ldots, 10$, $\theta_{m0} = .2$, and with the alternative values generated uniformly over $(.2, 1)$. The results are summarized in Table 3. Observe the differences in the sizes allocated by the optimal procedure and the Sidak procedure. The efficiency of the optimal procedure is $(7.4303/7.1846) \times 100 = 103.4\%$.

8.4. *A Size-Investing Strategy.* In each of the three concrete examples we have observed from the figures and tables the phenomenon where, among the M tests, those with low powers and those with high powers, or equivalently, those with small effect sizes and those with large effect sizes, are allocated

TABLE 3

Illustrative example for the multiple testing problem for Bernoulli/binomial distributions with $M = 10$ and $n_m = 5$ for $m = 1, 2, \ldots, 10$ and $\alpha = .05$. The null values are identically equal 0.2, whereas the alternative values were generated uniformly (sorted in the table).

m	θ_{m0}	θ_{m1}	Optimal Sizes	Optimal Powers	Sidak Sizes	Sidak Powers
1	0.2	0.3295	9.554e-04	0.0193	0.0051	0.0603
2	0.2	0.4407	1.043e-02	0.2774	0.0051	0.2078
3	0.2	0.5999	9.764e-03	0.6587	0.0051	0.5759
4	0.2	0.6610	6.371e-03	0.7751	0.0051	0.7224
5	0.2	0.6826	6.369e-03	0.8183	0.0051	0.7692
6	0.2	0.7807	6.405e-03	0.9527	0.0051	0.9276
7	0.2	0.7959	6.369e-03	0.9644	0.0051	0.9434
8	0.2	0.8438	3.562e-03	0.9652	0.0051	0.9779
9	0.2	0.9490	8.643e-04	0.9988	0.0051	0.9996
10	0.2	0.9992	4.198e-06	0.9999	0.0051	1.0000
Total	NA	NA	NA	7.4303	NA	7.1846

relatively small sizes in the FWER-controlling optimal procedure, for the overall FWER-size α used in these examples. The tests which are getting the larger sizes are those with moderate powers or moderate effect sizes. We refer to this as a size-investing strategy in the multiple hypotheses testing setting. The theoretical basis for this strategy, at least under the conditions of Theorem 5.1, is the first condition for optimality stated in that theorem which is given in (5.1). This condition is tied-in to the rate of change or derivative of the power as the size changes of the Neyman-Pearson MP test, but with this rate of change attenuated by the size.

This strategy can be also explained in an intuitive way. With the overall goal of getting more real discoveries while controlling the proportion of false discoveries for a pre-specified, usually small, overall size α, the optimal procedure dictates that not much size should be accorded those tests with either very low power or very high power. The former case will not lead to any discoveries anyway if the size that could be allocated is small, while the latter case will lead to discoveries even if the test sizes are made small. Thus, there is more to be gained by investing larger sizes on those tests that are of moderate power, and an appropriate tweaking of their test sizes, which is according to the condition in (5.1), improves the ability to achieve more real discoveries. However, this phenomenon is dependent on the magnitude of the overall size. If this overall size is made larger, then more leeway may then ensue to the extent that it may then be more beneficial to allocate more size also to those with low power since those tests with moderate powers when they had small sizes may now have larger powers because of the consequent increase in their sizes. From a technical standpoint, the precise and crucial determinant of where the differential sizes should be allocated are the rates of change of each of the power functions as size is changed, with some size-attenuation.

Interestingly, a tangential real-life application of this size-investing strategy occurred during the recent American presidential election, with the total resources (financial, manpower, etc.) available to the candidates analogous to the overall size in the multiple testing problem. In the waning days of the campaign, the major presidential candidates, then-Senator Barack Obama of the Democratic Party and Senator John McCain of the Republican Party, focussed their campaign efforts, in terms of allocating their financial and manpower resources, in the 'battleground states' of North Carolina, Virginia, and Pennsylvannia, while basically ignoring the 'in-the-bag states' of South Carolina, which was then expected to vote for McCain, and California, which was then expected to vote for Obama. It should also be noted that by virtue of the deep resources of the Obama campaign, it was able to allocate more resources even in states that traditionally voted Republican, whereas the McCain campaign, which had a relatively smaller war chest, had to 'drop' some states (e.g., Michigan) in their campaign. These opposing behaviors of the two camps could be explained by the size-investing strategy with proper accounting of each campaign's overall resources.

9. FDR-Controlling Procedures. In this section we use the FWER-controlling optimal procedure as an anchor to develop a procedure that strongly controls the (expected) false discovery rate. Thus, for a specified threshold $q^* \in (0,1)$, we will construct a $\delta^* \in \mathcal{D}$ such that $R_1(\delta^*, Q) \le q^*$ whatever the true probability measure Q of X is. Note that we are not restricting δ^* to be in $\mathcal{D}_0$, rather, we now allow the possibility that each component of δ^* may depend on the whole data X, instead of just X_m.

Before developing the decision function, we first recall the FDR-controlling procedure of (1), which we shall label as the δ^{BH} MHTDF. Denote by $(P_m(X_m), m \in \mathcal{M})$ the collection of p-value statistics associated with the M Neyman-Pearson tests, and denote by $P_{(1)} \le P_{(2)} \le \ldots \le P_{(M)}$ be the associated ordered p-value statistics. Note that for each $m \in \mathcal{M}$, $P_{(m)} = P_{(m)}(X)$, that is, the order statistics depend on the whole data X. Let $H_{(m)0}$ and $H_{(m)1}$ be the null and alternative hypotheses, respectively, associated with $P_{(m)}$. The basic assumption is that for $m \in \mathcal{M}_0(Q)$, P_m has a $U[0,1]$ distribution, aside from the independence condition in (2.3). Define the mapping $J : (\mathcal{X}, \mathcal{B}) \to (\{0, 1, \ldots, M, \infty\}, \sigma\{0, 1, \ldots, M, \infty\})$ according to

$$J^{BH}(x) = \max\left\{k \in \{0,1,\ldots,M\} : \ P_{(k)}(x) \le \frac{q^* k}{M}\right\} \tag{9.1}$$

with $P_{(0)}(x_m) = 0$ and $\max\emptyset = \infty$. Then $\delta^{BH} \in \mathcal{D}$ is the MHTDF which rejects all $H_{(m)0}$s with $m \in \{0,1,\ldots,J^{BH}(X)\}$ and accepts all $H_{(m)0}$s with $m \in \{J^{BH}(X)+1, J^{BH}(X)+2,\ldots,M\}$. In their seminal paper, (1) elegantly proved, using a proof by induction, that $R_1(\delta^{BH}, Q) \le q^*$ whatever is the true probability measure Q, so that δ^{BH} provides strong FDR-control at q^*.

Note that (31) presented an alternative proof, using reverse martingales, of the strong FDR-control by δ^{BH}. For our proposed FDR-controlling extension, we shall also employ a reverse martingale argument in establishing the main result.

We observe however for δ^{BH} that, just like the FWER-controlling procedures of Sidak and Bonferonni, it does not utilize the possibly differing powers of the M tests, hence there is the possibility of improving on δ^{BH} by exploiting the powers of the individual tests analogously to the development of the optimal FWER-controlling procedure.

To motivate and relate our proposed FDR-controlling procedure to δ^{BH}, we note the alternative description of δ^{BH} as presented in Section 3.3 of (1). Define $\alpha^{BH} : (\mathcal{X}, \mathcal{B}) \to ([0,1], \sigma[0,1])$ according to

$$\alpha^{BH} \equiv \alpha^{BH}(x; q^*) = \sup\left\{\alpha \in [0,1] : \ \alpha \le \frac{q^*}{M} \sum_{m\in\mathcal{M}} I\{P_m(X_m) \le \alpha\}\right\}. \tag{9.2}$$

The mth component of δ^{BH} can now be equivalently defined via $\delta^{BH}_m(x) = I\{P_m(x_m) \le \alpha^{BH}(x; q^*)\}$ for all $m \in \mathcal{M}$.

Let us now denote by

$$\alpha \mapsto \eta(\alpha) = (\eta_m(\alpha),\ m \in \mathcal{M})$$

the mapping which produces the optimal size vector for FWER control at α where $\alpha \in [0,1]$. The existence of such a mapping is guaranteed by Theorem 4.1. Suppose that it is desired to strongly control the FDR at $q^* \in (0,1)$. Let us define $\alpha^* : (\mathcal{X} \times [0,1]^M, \mathcal{B} \otimes \sigma\{[0,1]^M\}) \to ([0,1], \sigma[0,1])$ according to

$$\begin{aligned} \alpha^* &\equiv \alpha^*(x, u; q^*) \\ &= \sup\left\{\alpha \in [0,1] : \sum_{m\in\mathcal{M}} \eta_m(\alpha) \le q^* \sum_{m\in\mathcal{M}} \delta_m(x_m, u_m; \eta_m(\alpha))\right\}. \end{aligned} \tag{9.3}$$

Our proposed procedure for controlling the FDR at q^* is the MHTDF $\delta^* \in \mathcal{D}$ given by

$$\delta^*(x,u) = (\delta_m(x_m, u_m; \eta_m(\alpha^*(x,u;q^*))),\ m \in \mathcal{M}), \tag{9.4}$$

where we note that the $\delta_m(x_m, u_m; \eta_m)$ is the size-η_m Neyman-Pearson MP test of H_{m0} versus H_{m1}. It should be pointed out that the mth component δ^*_m of δ^* depends on the whole data (x,u), not just on (x_m, u_m). This is by virtue of the fact that α^* depends on the whole data (x,u). This is analogous to the situation with the (1) procedure where, even though for each $m \in \mathcal{M}$, the (initial) p_m-value is computed based only on x_m, the final FDR-controlling procedure δ^{BH} has all its components δ^{BH}_m, $m \in \mathcal{M}$, depending on the whole data x. Notice also that, whereas the observable data X needed to implement δ^{BH} needs to be of a continuous-type, else the uniformity of the p_m-value statistic for $m \in \mathcal{M}_0(Q)$ breaks down, in the proposed extension such a restriction need not be imposed since the MP tests allow both discrete and continuous observables.

We establish that the MHTDF δ^* in (9.4) provides strong FDR control at q^*.

THEOREM 9.1. *Let Q be the true, but unknown, probability measure of X, and assume that for each $m \in \mathcal{M}$, the Neyman-Pearson MP test of $H_{m0} : Q_m = Q_{m0}$ versus $H_{m1} : Q_m = Q_{m1}$ of level η_m satisfies*

$$E_{X_m \sim Q_{m0}}[\delta_m(X_m, U_m; \eta_m)] = \eta_m$$

for each $\eta_m \in (0,1)$. If the collection of optimal FWER-controlling size vectors $\{\eta(\alpha) = (\eta_m(\alpha), m \in \mathcal{M}) :\ \alpha \in (0,1)\}$ satisfies the conditions (i) for every $m \in \mathcal{M}$, the mapping $\alpha \mapsto \eta_m(\alpha)$ is nondecreasing; and (ii) for every $\alpha \in (0,1)$,

$$\frac{(M-1)\max_{m\in\mathcal{M}} \eta_m(\alpha)}{\sum_{m\in\mathcal{M}} \eta_m(\alpha)} \le 1, \tag{9.5}$$

then the MHTDF $\delta^* \in \mathcal{D}$ given in (9.4) strongly controls the false discovery rate (FDR) at q^* under Q, that is, whatever the true Q is, $R_1(\delta^*, Q) \le q^*$.

PROOF. We first note that for every $\alpha \in [0,1)$, $\eta_m(\alpha) \le \alpha$ for all $m \in \mathcal{M}$. This follows from the inequality $\eta_m(\alpha) \le \alpha$. For each $\alpha \in (0,1]$, define the sigma-field

$$\mathcal{F}_\alpha = \sigma\{\delta_m(X_m, U_m; \eta_m(\beta)) :\ m \in \mathcal{M}, \alpha \le \beta < 1\},$$

and let $\mathcal{F}_0 = \bigvee_{\alpha \in (0,1]} \mathcal{F}_\alpha$. Observe that for $0 \le \alpha < \beta \le 1$, $\mathcal{F}_\alpha \supset \mathcal{F}_\beta$. Denote by $\mathbf{F} = \{\mathcal{F}_\alpha : \alpha \in [0,1]\}$ the induced filtration. Henceforth, for conciseness, we shall drop (X_m, U_m) in $\delta_m(X_m, U_m; \eta_m)$ and simply write it as $\delta_m(\eta_m)$. Define the $\mathbf{F}$-adapted processes, indexed by $\alpha \in [0,1]$, by

$$T_0(\alpha) = \sum_{m \in \mathcal{M}_0} \delta_m(\eta_m(\alpha)) \quad \text{and} \quad T(\alpha) = \sum_{m \in \mathcal{M}} \delta_m(\eta_m(\alpha)).$$

Then the FDR of the MHTDF δ^* defined in (9.4) is

$$\begin{aligned} R_1(\delta^*, Q) &= E_{X \sim Q}\left\{\frac{T_0(\alpha^*)}{T(\alpha^*)} I\{T(\alpha^*) > 0\}\right\} \\ &= E\left[E_{X \sim Q}\left\{\frac{T_0(\alpha^*)}{T(\alpha^*)} I\{T(\alpha^*) > 0\} | \mathcal{M}_0\right\}\right], \end{aligned}$$

where we note that, given Q, $\mathcal{M}_0 = \mathcal{M}_0(Q)$ is a random, albeit degenerate, subset of $\mathcal{M}$. Thus, the outer expectation is an expectation with respect to this degenerate probability measure. Let us focus on the inner expectation. If $M_0 = 0$, then clearly, this expectation is zero since $T_0(\alpha^*) = 0$, so that trivially this expectation is bounded above by q^*.

Next we consider the case where $M_0 \in \{1, 2, \ldots, M-1\}$. Observe from the definition of $\alpha^* = \alpha^*(q^*)$ given in (9.3) that this is a $\mathbf{F}$-stopping time. We also observe that when evaluated at $\alpha = \alpha^*$, we have the inequality

$$T(\alpha^*) \ge \frac{\sum_{m \in \mathcal{M}} \eta_m(\alpha^*)}{q^*} \equiv \frac{\eta_\bullet(\alpha^*)}{q^*}.$$

Consequently,

$$\begin{aligned} & E_{X \sim Q}\left\{\frac{T_0(\alpha^*)}{T(\alpha^*)} I\{T(\alpha^*) > 0\} | \mathcal{M}_0\right\} \\ &\quad \le q^* E_{X \sim Q}\left\{\frac{T_0(\alpha^*)}{\eta_\bullet(\alpha^*)} I\{T(\alpha^*) > 0\} | \mathcal{M}_0\right\} \\ &\quad = q^* E_{X \sim Q}\left\{\frac{T_0(\alpha^*)}{\eta_\bullet(\alpha^*)} | \mathcal{M}_0\right\}, \end{aligned} \tag{9.6}$$

the last equality following since $T_0^*(\alpha) I\{T(\alpha) > 0\} = T_0^*(\alpha)$ for every $\alpha \in [0,1]$. Next, consider the $\mathbf{F}$-adapted process $\{T_0^*(\alpha) : \alpha \in (0,1]\}$ with

$$T_0^*(\alpha) = \sum_{m \in \mathcal{M}_0} \frac{\delta_m(\eta_m(\alpha))}{\eta_m(\alpha)}.$$

We have for $0 < \alpha < \beta \leq 1$ that

$$
\begin{aligned}
E_{X\sim Q}\{T_0^*(\alpha)|\mathcal{F}_\beta, \mathcal{M}_0\} &= E_{X\sim Q}\left\{\sum_{m\in\mathcal{M}_0} \frac{\delta_m(\eta_m(\alpha))}{\eta_m(\alpha)}|\mathcal{F}_\beta, \mathcal{M}_0\right\} \\
&= \sum_{m\in\mathcal{M}_0} E_{X\sim Q}\left\{\frac{\delta_m(\eta_m(\alpha))}{\eta_m(\alpha)}|\mathcal{F}_\beta, \mathcal{M}_0\right\} \\
&= \sum_{m\in\mathcal{M}_0} \delta_m(\eta_m(\beta)) E_{X\sim Q}\left\{\frac{\delta_m(\eta_m(\alpha))}{\eta_m(\alpha)}|\delta_m(\eta_m(\beta)) = 1, \mathcal{F}_\beta, \mathcal{M}_0\right\} \\
&= \sum_{m\in\mathcal{M}_0} \delta_m(\eta_m(\beta)) E_{X_m\sim Q_{m0}}\left\{\frac{\delta_m(\eta_m(\alpha))}{\eta_m(\alpha)}|\delta_m(\eta_m(\beta)) = 1\right\} \\
&= \sum_{m\in\mathcal{M}_0} \frac{\delta_m(\eta_m(\beta))}{\eta_m(\alpha)} \frac{\mathbf{P}\{\delta_m(\eta_m(\alpha)) = 1|X_m \sim Q_{m0}\}}{\mathbf{P}\{\delta_m(\eta_m(\beta)) = 1|X_m \sim Q_{m0}\}} \\
&= \sum_{m\in\mathcal{M}_0} \frac{\delta_m(\eta_m(\beta))}{\eta_m(\alpha)} \frac{\eta_m(\alpha)}{\eta_m(\beta)} \\
&= \sum_{m\in\mathcal{M}_0} \frac{\delta_m(\eta_m(\beta))}{\eta_m(\beta)} \\
&= T_0^*(\beta),
\end{aligned}
$$

where the fourth equality follows from the assumed independence condition given by (2.3). The fifth equality follows from the assumed nondecreasing property of the mapping $\alpha \mapsto \eta_m(\alpha)$, which implies that for $\alpha < \beta$, $\{\delta_m(\eta_m(\alpha)) = 1\} \Rightarrow \{\delta_m(\eta_m(\beta)) = 1\}$ and the condition that $\mathbf{P}\{\delta_m(\eta_m) = 1|X_m \sim Q_{m0}\} = E_{X_m\sim Q_{m0}}[\delta_m(\eta_m)] = \eta_m$. This sequence of equalities establishes the result that $\{(T_0^*(\alpha), \mathcal{F}_\alpha) : \alpha \in (0, 1]\}$ is a reverse martingale process. Define

$$T_0^*(0) = \liminf_{\alpha\in(0,1]} T_0^*(\alpha) = \limsup_{\alpha\in(0,1]} T_0^*(\alpha),$$

which, by Doob's martingale convergence theorem, is well-defined. Then the extended collection $\{(T_0^*(\alpha), \mathcal{F}_\alpha) : \alpha \in [0, \alpha_U]\}$ is a reverse martingale process.

Going back to the expectation portion of the upper bound given in (9.6), we have

$$E_{X\sim Q}\left\{\frac{T_0(\alpha^*)}{\eta_\bullet(\alpha^*)}|\mathcal{M}_0\right\} = E_{X\sim Q}\left\{\frac{\sum_{m\in\mathcal{M}_0}\delta_m(\eta_m(\alpha^*))}{\eta_\bullet(\alpha^*)}|\mathcal{M}_0\right\}$$

$$\begin{aligned}
&= E_{X\sim Q}\left\{\sum_{m\in\mathcal{M}_0}\frac{\delta_m(\eta_m(\alpha^*))}{\eta_m(\alpha^*)}\frac{\eta_m(\alpha^*)}{\eta_\bullet(\alpha^*)}|\mathcal{M}_0\right\}\\
&\le \left[\sup_{\alpha\in(0,1]}\frac{\max_{m\in\mathcal{M}_0}\eta_m(\alpha)}{\eta_\bullet(\alpha)}\right]E_{X\sim Q}\{T_0^*(\alpha^*)|\mathcal{M}_0\}\\
&\le \frac{1}{M-1}E_{X\sim Q}\{T_0^*(\alpha^*)|\mathcal{M}_0\}\\
&= \frac{1}{M-1}E_{X\sim Q}\{T_0^*(1)|\mathcal{M}_0\}\\
&= \frac{1}{M-1}\sum_{m\in\mathcal{M}_0}\frac{E_{X_m\sim Q_{m0}}\{\delta_m(\eta_m(1))|\mathcal{M}_0\}}{\eta_m(1)}\\
&= \frac{1}{M-1}\sum_{m\in\mathcal{M}_0}\frac{\eta_m(\alpha_U)}{\eta_m(\alpha_U)}\\
&= \frac{M_0}{M-1}\\
&\le 1,
\end{aligned}$$

where we used condition (9.5) in the statement of the theorem to get the second inequality, and the Optional Sampling Theorem for martingales (cf., (2)) to get the third equality, and the last inequality since we are in the case with $M_0 \le M-1$. Thus, we have established that, whatever $\mathcal{M}_0$ is provided that $|\mathcal{M}_0| \le M-1$, the inequality

$$E_{X\sim Q}\left\{\frac{T_0(\alpha^*)}{T(\alpha^*)}I\{T(\alpha^*)>0\}|\mathcal{M}_0\right\} \le q^*.$$

Taking expectation with respect to $\mathcal{M}_0$, which as pointed out earlier is really a degenerate probability measure, yields the result that $R_1(\delta^*, Q) \le q^*$.

Finally, we deal with the case when $M_0 = M$, that is, when all the H_{m0}s are correct. Recall the Sidak FWER-α controlling sizes given by

$$\eta_m^S(\alpha) = 1-(1-\alpha)^{1/M},\ m\in\mathcal{M}.$$

The vector $\eta^S(\alpha) = (\eta_m^S(\alpha):\ m\in\mathcal{M})$ clearly satisfies (9.5). Define

$$\alpha^S = \alpha^S(q^*) =$$

$$\sup\left\{\alpha\in[0,1]:\ \eta_\bullet^S(\alpha) = M[1-(1-\alpha)^{1/M}] \le q^*\sum_{m\in\mathcal{M}}\delta_m(\eta_m^S(\alpha))\right\} \tag{9.7}$$

and the associated MHTDF given by $\delta^S = (\delta_m(\eta_m^S(\alpha)) : \ m \in \mathcal{M})$. Since in the proof above for the case with $M_0 \in \{0, 1, 2, \ldots, M-1\}$ it is not necessary that the size vector $(\eta_m(\alpha) : \ m \in \mathcal{M})$ be the FWER-controlling optimal size vector, then the proof also holds when we use the Sidak size vector. Furthermore, since the Sidak size vector satisfies condition (9.5) even when $M_0 = M$, then for the Sidak sizes, we have

$$R_1(\delta^S, Q) \leq q^* \text{ for all } Q \text{ including } Q_0.$$

Define for every $\alpha \in [0, 1]$ the processes

$$V^*(\alpha) = \frac{\sum_{m \in \mathcal{M}} \delta_m(\eta_m(\alpha))}{\sum_{m \in \mathcal{M}} \eta_m(\alpha)} \quad \text{and} \quad V^S(\alpha) = \frac{\sum_{m \in \mathcal{M}} \delta_m(\eta_m^S(\alpha))}{\sum_{m \in \mathcal{M}} \eta_m^S(\alpha)}$$

with $V^*(0) = V^S(0) = 0$. Observe that, under Q_0, the expectations of $V^*(\alpha)$ and $V^S(\alpha)$ are both equal to 1 for each $\alpha \in (0, 1]$. We may re-express both α^* and α^S via

$$\begin{aligned} \alpha^* &= \sup\{\alpha \in [0, 1] : \ V^*(\alpha) \geq 1/q^*\}; \\ \alpha^S &= \sup\{\alpha \in [0, 1] : \ V^S(\alpha) \geq 1/q^*\}. \end{aligned}$$

Since $q^* < 1$, by Lemma 9.1 established below, it follows that

$$\mathbf{P}_{X \sim Q_0}\left\{V^*(\alpha) \geq \frac{1}{q^*}\right\} \leq \mathbf{P}_{X \sim Q_0}\left\{V^S(\alpha) \geq \frac{1}{q^*}\right\}.$$

This implies that, under Q_0, $\alpha^* \overset{st}{\leq} \alpha^S$. From this it follows that

$$R_1(\delta^*, Q_0) = \mathbf{P}_{X \sim Q_0}(\alpha^* > 0) \leq P_{X \sim Q_0}(\alpha^S > 0) = R_1(\delta^S, Q_0) \leq q^*.$$

Though it is not essential in the proof of our result, notice that, in fact, $R_1(\delta^S, Q_0) = q^*$. This can be seen by examining the proof above and seeing where the inequalities are, which will all turn out to be equalities for the Sidak size vector and associated MHTDF. We have thus completed the proof that, whatever Q is, $R_1(\delta^*, Q) \leq q^*$, hence our theorem. ☐

We now establish the lemma that was used in the preceding proof.

LEMMA 9.1. *Let $W_1, W_2, \ldots, W_M$ be independent random variables with $W_m \sim Ber(\eta_m)$ where $\eta = (\eta_1, \eta_2, \ldots, \eta_M) \in [0, 1]^M$. For $a \geq 0$, define*

$$h_a(\eta) = \mathbf{P}\left\{\frac{\sum_{m=1}^M W_m}{\sum_{m=1}^M \eta_m} \geq a\right\}.$$

Letting $UB(C_\alpha) = \{\eta \in [0,1]^M : \prod_{m=1}^M (1-\eta_m) = 1-\alpha\}$ *for* $\alpha \in (0,1)$, *then, for every* $a \geq 1$, *we have*

$$\sup\{h_a(\eta) : \eta \in UB(C_\alpha)\} = h_a(\eta^S(\alpha)),$$

where $\eta_m^S(\alpha) = 1-(1-\alpha)^{1/M}$ *for all* m, *the Sidak sizes.*

PROOF. Let $Z_1, Z_2, \ldots, Z_M$ be independent random variables with $Z_m \sim Ber(p_m)$ and denote by $\bar{p} = \frac{1}{M}\sum_{m=1}^M p_m$. For $t \geq 0$, let

$$h_t^*(p_1, p_2, \ldots, p_M) = \mathbf{P}\left\{\sum_{m=1}^M Z_m \geq t\right\}.$$

In (12) (see also pages 375–376 of (17)) proved that if $M\bar{p} \leq t \leq M$, then $h_t^*(p_1, p_2, \ldots, p_M) \leq h_t^*(\bar{p}, \bar{p}, \ldots, \bar{p})$. In the setting of the lemma, define

$$p_m = p_m(\eta) = -\log(1-\eta_m),\ m = 1, 2, \ldots, M.$$

Then, $\eta \in UB(C_\alpha)$ iff $\sum_{m=1}^M p_m(\eta_m) = -\log(1-\alpha)$. For $a \geq 1$, we are then able to apply the result in (12) to conclude that for all $\eta \in UB(C_\alpha)$,

$$\begin{aligned} h_a(\eta) &= h^*_{-\frac{a}{M}\log(1-\alpha)}(p_1(\eta_1), \ldots, p_m(\eta_m)) \\ &\leq h^*_{-\frac{a}{M}\log(1-\alpha)}\left(-\frac{1}{M}\log(1-\alpha), \ldots, -\frac{1}{M}\log(1-\alpha)\right) \\ &= h_a\left(1-(1-\alpha)^{1/M}, \ldots, 1-(1-\alpha)^{1/M}\right). \end{aligned}$$

This establishes the lemma. □

COROLLARY 9.1. *If the conditions of Theorem 5.1 and condition (9.5) are satisfied, then the conclusion of Theorem 9.1 holds.*

PROOF. This immediately follows from Theorem 9.1 and Proposition 5.1 since the latter guarantees that the mappings $\alpha \mapsto \eta_m(\alpha)$ for $m \in \mathcal{M}$ are nondecreasing. □

We remark that the existence (and uniqueness) of an FWER-controlling optimal size vector did not require the differentiability conditions of Theorem 5.1. Thus, it is possible that the nondecreasing property of the mappings $\alpha \mapsto \eta_m(\alpha)$ for each $m \in \mathcal{M}$ still holds under weaker conditions than the differentiabiliy conditions in Theorem 5.1 and Proposition 5.1. In this regard we are still seeking a more general version of Proposition 5.1.

The next corollary establishes that the MHTDF δ^{BH} is a special case of the proposed FDR-controlling extension in (9.4).

COROLLARY 9.2. *If, for all $m \in \mathcal{M}, \pi_m(\eta) = \pi(\eta)$ for some power function $\eta \mapsto \pi(\eta)$, $\pi(\cdot)$ satisfies the differentiability conditions of Theorem 5.1, and the p-value statistic $P_m(X_m)$ associated with δ_m has a continuous distribution under Q_{m0}, then the MHTDF δ^{BH} is obtained as a special case of δ^* in (9.4) and also coincides with δ^S.*

PROOF. Since the power functions $\eta_m \mapsto \pi_m(\eta_m)$ are identical, this implies that $\forall m \in \mathcal{M}, \eta_m(\alpha) = \eta(\alpha)$ for some $\eta(\alpha)$. The definition of α^* in (9.3) becomes

$$\alpha^* = \sup\left\{\alpha \in [0,1]:\ M\eta(\alpha) \leq q^* \sum_{m\in\mathcal{M}} \delta_m(\eta(\alpha))\right\}. \tag{9.8}$$

Relabeling $\eta(\alpha)$ by just α in (9.8) and noting that each $\delta_m(\alpha)$ could be re-expressed via $\delta_m(\alpha) = I\{P_m(X_m) \leq \alpha\}$, where the continuity of the distribution of $P_m(X_m)$ under Q_{m0} allows the nonrandomized form of δ_m in terms of $P_m(X_m)$, then (9.8) is precisely the definition of α^{BH} in (9.2) in the alternative form of δ^{BH}. This also coincides with δ^S, so that that the martingale proof in Theorem 9.1 carries over to establishing, in an alternative manner, that the MHTDF δ^{BH} provides strong FDR control for all Q (see also the martingale proof of (31)). □

We provide remarks concerning the condition on the FWER-controlling optimal size vector in (9.5). As pointed out when dealing with the Sidak sizes and test and also in Corollary 9.2, (9.5) is trivially satisfied when all the mappings $\eta_m \mapsto \pi_m(\eta_m)$ are all identical. In the general case where these mappings are non-identical, condition (9.5) induces some form of control of their potential differences. For instance, the condition precludes the situation where $\eta_{m_0}(\alpha) = \alpha$ for some $m_0 \in \mathcal{M}_0$ and all other $\eta_m(\alpha) = 0$ for $m \neq m_0$. It is easy to verify that a sufficient, but not a necessary, condition in terms of the extreme values of the $\eta_m(\alpha)$s for (9.5) to hold is

$$\frac{\min_{m\in\mathcal{M}} \eta_m(\alpha)}{\max_{m\in\mathcal{M}} \eta_m(\alpha)} \geq \frac{M-2}{M-1} \text{ for all } \alpha \in (0,1), \tag{9.9}$$

which is a very strict condition especially when M becomes large. We surmise that a weaker condition is possible without destroying the strong FDR-control property of δ^*; however, this seems to require an alternative proof of the theorem which does not utilize a martingale approach. This conjecture that the FDR-control still holds under a weaker condition seems to be borne out by the results of the simulation studies in Section 10 where the generation of the effect sizes do not guarantee that condition (9.5) will always hold.

Finally, we point out that the MHTDF $\delta^*(x,u)$ could also be described in a form akin to that of $\delta^{BH}(x)$ involving $J^{BH}(x)$ in (9.1). For this purpose, define $\alpha^*(x,u) = (\alpha_m^*(x,u), m \in \mathcal{M})$ via

$$\alpha_m^* = \alpha_m^*(x,u) = \inf\{\alpha \in [0,1]:\ \delta_m(x_m, u_m; \eta_m(x,u;\alpha)) = 1\}. \tag{9.10}$$

The value $\alpha_m^*(x,u)$ can be interpreted as the smallest FWER-size such that H_{m0} is rejected for the data (x,u) in light of all the other hypotheses in the multiple testing problem. Denote by $(\alpha_{(m)}^*:\ m \in \mathcal{M})$ the ordered values of $(\alpha_m^*:\ m \in \mathcal{M})$, and let $H_{(m)0}$ and $H_{(m)1}$ the null and alternative hypotheses associated with $\alpha_{(m)}^*$. We may now define

$$J^* = J^*(x,u) = \max\left\{k \in \mathcal{M}:\ \sum_{m\in\mathcal{M}} \eta_m(\alpha_{(k)}^*(x,u)) \le q^* k\right\}. \tag{9.11}$$

Then the MHTDF $\delta^*(X,U)$ rejects $H_{(m)0}$ for $m \in \{1,2,\ldots,J^*(X,U)\}$ and accepts $H_{(m)0}$ for $m \in \{J^*(X,U)+1, J^*(X,U)+2,\ldots,M\}$. In our `R` implementation of δ^*, we find that the program code which calculates α^* is computationally more efficient than the program code which uses J^*, which is in contrast to that of δ^{BH} which is easier to implement via the J^{BH}-form.

It is interesting to note the distribution of $\alpha_{(1)}$ under $Q = Q_0$. We have, for $a \in (0,1)$,

$$\begin{aligned}\mathbf{P}_{X\sim Q_0}(\alpha_{(1)} > a) &= \mathbf{P}_{X\sim Q_0}\left\{\bigcap_{m\in\mathcal{M}} [\delta_m(\eta_m(a)) = 0]\right\} \\ &= \prod_{m\in\mathcal{M}} [1-\eta_m(a)] = 1-a,\end{aligned} \tag{9.12}$$

where we used the independence condition under $Q = Q_0$, so that $\alpha_{(1)}$ has a standard uniform distribution when all the null hypotheses are correct! Using this result and the following lemma concerning lower and upper bounds of $\eta_\bullet$ for $\eta \in UB(C_\alpha)$, we obtain a lower bound of $R_1(\delta^*, Q_0)$, the false discovery rate when all the null hypotheses are correct, in Proposition 9.1.

LEMMA 9.2. *Every $\eta \in UB(C_\alpha)$ satisfies*

$$\alpha \le \eta_\bullet = \sum_{m\in\mathcal{M}} \eta_m \le \min\left\{-\log(1-\alpha), M[1-(1-\alpha)^{1/M}]\right\}.$$

PROOF. For $\eta \in UB(C_\alpha)$, let $W_1, W_2, \ldots, W_M$ be independent Bernoulli random variables with $W_m \sim Ber(\eta_m)$. Then, using Bonferonni's inequality, we obtain

$$\alpha = 1 - \prod_{m\in\mathcal{M}} [1-\eta_m] = \mathbf{P}\left\{\bigcup_{m\in\mathcal{M}} [W_m = 1]\right\}$$
$$\leq \sum_{m\in\mathcal{M}} \mathbf{P}\{W_m = 1\} = \sum_{m\in\mathcal{M}} \eta_m = \eta_\bullet,$$

thereby proving the left-hand inequality.

Since for every $a \in [0,1)$, $-\log(1-a) \geq a$, then from the constraint condition $\sum_{m\in\mathcal{M}} \log(1-\eta_m) = \log(1-\alpha)$, we obtain $\eta_\bullet \leq -\log(1-\alpha)$. But also, since the mapping $a \mapsto \log(1-a)$ is concave in $[0,1)$, then

$$\log\{(1-\alpha)^{1/M}\} = \frac{1}{M}\sum_{m\in\mathcal{M}} \log(1-\eta_m) \leq \log\left(1 - \frac{1}{M}\sum_{m\in\mathcal{M}} \eta_m\right).$$

This implies that $(1-\alpha)^{1/M} \leq 1 - \eta_\bullet/M$ which is equivalent to $\eta_\bullet \leq M[1-(1-\alpha)^{1/M}]$. This completes the proof of the lemma. □

PROPOSITION 9.1. *For $q^* \in (0,1)$, we have*

$$\max\left\{1-\exp(-q^*), 1-\left(1-\frac{q^*}{M}\right)^M\right\} \leq R_1(\delta^*, Q_0) \leq q^*.$$

PROOF. It only remains to show the left-hand inequality. We have that

$$\begin{aligned} R_1(\delta^*, Q_0) &= P_{X\sim Q_0}(\alpha^* > 0) \\ &= \mathbf{P}_{X\sim Q_0}\left\{\bigcup_{m=1}^{M}\left[\eta_\bullet(\alpha_{(m)}) \leq mq^*\right]\right\} \\ &\geq \mathbf{P}_{X\sim Q_0}\{\eta_\bullet(\alpha_{(1)}) \leq q^*\}. \end{aligned}$$

We now use the upper bound for $\eta_\bullet$ provided in Lemma 9.2. We have from the first component in this upper bound that $\mathbf{P}_{X\sim Q_0}\{\eta_\bullet(\alpha_{(1)}) \leq q^*\} \geq \mathbf{P}_{X\sim Q_0}\{-\log(1-\alpha_{(1)}) \leq q^*\} = 1-\exp(-q^*)$ since $-\log(1-\alpha_{(1)})$ has a unit exponential distribution owing to the standard uniformity of $\alpha_{(1)}$ under Q_0. But by using the second component in the upper bound, we also have

$$\begin{aligned} \mathbf{P}_{X\sim Q_0}\{\eta_\bullet(\alpha_{(1)}) \leq q^*\} &\geq \mathbf{P}_{X\sim Q_0}\{M[1-(1-\alpha_{(1)})^{1/M}] \leq q^*\} \\ &= \mathbf{P}_{X\sim Q_0}\left\{\alpha_{(1)} \leq 1-\left(1-\frac{q^*}{M}\right)^M\right\} \\ &= 1-\left(1-\frac{q^*}{M}\right)^M. \end{aligned}$$

These results establish the lower bound in the proposition. □

10. A Modest Simulation. In this section we present the results of a modest simulation study comparing the performance of the MHTDF δ^* and MHTDF δ^{BH} in terms of FDR and MDR. More elaborate simulated comparisons will be presented in a separate paper. The results presented here are limited to demonstrating numerically, in a specific Gaussian model, that δ^* achieves the desired FDR-control, as is the MHTDF δ^{BH}, and that δ^* also achieves a lower MDR relative to δ^{BH}.

The simulation model is similar to the first example illustrating the optimal FWER-controlling procedure. In this model, for each $m \in \mathcal{M}$, the observables are $X_m \sim N(\mu_m, 1)$ which are independently generated. The mth pair of hypotheses is $H_{m0} : \mu_m \leq 0$ versus $H_{m1} : \mu_m > 0$ with UMP size-η_m test of form $\delta_m(X_m; \eta_m) = I\{X_m > \Phi^{-1}(1 - \eta_m)\}$. The true values of the means μ_ms are $\mu_m = \xi_m \theta_m, m \in \mathcal{M}$, with $\theta_m \sim Ber(p)$ and effect sizes $\xi_m \sim |N(\nu, 1)|$, which are again independently generated from each other. In the simulation, the parameter combinations were induced by taking the number of pairs of hypotheses $M \in \{20, 50, 100\}$, the proportion of true alternative hypotheses $p \in \{.1, .2, .4\}$, and the mean of the effect size-generating normal distribution $\nu \in \{1, 2, 4\}$. In implementing the MHTDFs δ^* and δ^{BH}, we utilized an FDR-threshold of $q^* \in \{.05, .10\}$. Because the computational implementation of the MHTDF δ^* takes a relatively long time, for each of the combinations of (q^*, M, ν, p), we only replicated the basic experiment 200 times. For each simulation parameter combination, the simulated FDR and MDR were the averages of the observed FDR and MDR over the 200 replications.

Tables 4 and 5 present the results of this modest simulation study for $q^* = .05$ and $q^* = .10$, respectively. From these tables, we observe that both δ^* and δ^{BH} fulfill the FDR-constraint, and in fact this happens in a generally conservative fashion, which is as predicted by theory. More importantly, the MDR-performance of δ^* is better compared to that of δ^{BH} and this dominance holds true for all the simulation parameter combinations that were considered. Observe that as M is increased with (ν, p) remaining the same, there is an increase in their MDRs; whereas, when ν is increased, which has the effect of increasing the effect sizes, their MDRs decrease. Interestingly, the impact of a change of value in p, the proportion of true alternative hypotheses, did not necessarily translate into a monotone change in their MDRs, especially when $M = 20$, though for the larger M-values, the MDR change appears to be monotonically decreasing.

11. Concluding Remarks. This paper provides a modest resolution on the issue of the use of the individual power functions in multiple hy-

TABLE 4

Comparison of the false discovery rate (FDR) and missed discovery rate (MDR) performance of MHTDF δ^* and δ^{BH} under a variety of simulation parameters. This table is for $q^* = .05$. The FDR and MDR are in percentages. The number of replications is 200.

	q^*	M	ν	p	δ^*-FDR	δ^*-MDR	δ^{BH}-FDR	δ^{BH}-MDR
1	0.05	20.00	1.00	0.10	0.92	73.92	4.67	75.17
2	0.05	20.00	1.00	0.20	5.04	84.80	4.04	87.56
3	0.05	20.00	1.00	0.40	3.38	82.54	2.77	85.38
4	0.05	20.00	2.00	0.10	4.27	54.62	5.29	57.27
5	0.05	20.00	2.00	0.20	3.39	64.23	4.17	65.06
6	0.05	20.00	2.00	0.40	3.13	60.71	3.97	62.09
7	0.05	20.00	4.00	0.10	4.25	15.51	5.86	15.82
8	0.05	20.00	4.00	0.20	3.16	11.97	2.46	12.05
9	0.05	20.00	4.00	0.40	3.59	7.91	3.82	9.50
10	0.05	50.00	1.00	0.10	3.92	87.96	4.92	90.23
11	0.05	50.00	1.00	0.20	1.89	88.22	3.26	90.71
12	0.05	50.00	1.00	0.40	3.64	84.87	3.57	88.10
13	0.05	50.00	2.00	0.10	3.85	69.86	3.75	70.37
14	0.05	50.00	2.00	0.20	5.40	67.33	4.96	69.22
15	0.05	50.00	2.00	0.40	3.18	60.00	2.73	60.80
16	0.05	50.00	4.00	0.10	4.93	15.04	4.30	16.35
17	0.05	50.00	4.00	0.20	5.04	9.95	4.52	13.04
18	0.05	50.00	4.00	0.40	2.76	7.33	3.11	9.76
19	0.05	100.00	1.00	0.10	2.52	90.12	1.71	92.28
20	0.05	100.00	1.00	0.20	3.93	89.44	4.24	93.19
21	0.05	100.00	1.00	0.40	3.07	86.22	2.63	89.60
22	0.05	100.00	2.00	0.10	4.52	73.44	4.64	75.26
23	0.05	100.00	2.00	0.20	4.06	67.65	3.47	69.23
24	0.05	100.00	2.00	0.40	2.87	58.30	2.77	59.82
25	0.05	100.00	4.00	0.10	4.30	16.34	4.54	18.10
26	0.05	100.00	4.00	0.20	4.16	10.34	4.74	12.09
27	0.05	100.00	4.00	0.40	3.07	6.31	2.93	8.32

Table 5

Comparison of the false discovery rate (FDR) and missed discovery rate (MDR) performance of MHTDF δ^* and δ^{BH} under a variety of simulation parameters. This table is for $q^* = .10$. The FDR and MDR are in percentages. The number of replications is 200.

	q^*	M	ν	p	δ^*-FDR	δ^*-MDR	δ^{BH}-FDR	δ^{BH}-MDR
1	0.10	20.00	1.00	0.10	4.75	69.94	8.25	71.65
2	0.10	20.00	1.00	0.20	12.88	80.88	8.04	82.70
3	0.10	20.00	1.00	0.40	7.14	75.23	7.38	79.35
4	0.10	20.00	2.00	0.10	9.44	48.52	9.72	49.60
5	0.10	20.00	2.00	0.20	7.94	55.16	8.25	56.34
6	0.10	20.00	2.00	0.40	5.74	51.81	5.16	52.79
7	0.10	20.00	4.00	0.10	9.96	11.27	9.40	12.93
8	0.10	20.00	4.00	0.20	7.18	5.54	7.00	9.41
9	0.10	20.00	4.00	0.40	7.30	4.66	6.79	6.88
10	0.10	50.00	1.00	0.10	7.78	84.34	8.49	86.55
11	0.10	50.00	1.00	0.20	5.34	83.67	6.78	86.20
12	0.10	50.00	1.00	0.40	6.29	77.91	7.15	81.17
13	0.10	50.00	2.00	0.10	7.91	62.58	9.25	64.10
14	0.10	50.00	2.00	0.20	9.30	58.18	8.60	61.21
15	0.10	50.00	2.00	0.40	5.39	48.68	5.91	50.79
16	0.10	50.00	4.00	0.10	8.85	10.76	9.04	11.21
17	0.10	50.00	4.00	0.20	8.72	6.51	8.28	8.06
18	0.10	50.00	4.00	0.40	5.57	4.12	6.24	6.76
19	0.10	100.00	1.00	0.10	6.28	87.29	5.04	89.91
20	0.10	100.00	1.00	0.20	7.22	85.05	9.29	89.05
21	0.10	100.00	1.00	0.40	6.02	79.29	6.84	83.33
22	0.10	100.00	2.00	0.10	9.86	65.92	9.67	67.70
23	0.10	100.00	2.00	0.20	8.12	59.17	8.53	60.34
24	0.10	100.00	2.00	0.40	5.84	48.14	5.98	49.64
25	0.10	100.00	4.00	0.10	8.30	11.48	9.17	13.93
26	0.10	100.00	4.00	0.20	8.49	6.21	8.30	8.49
27	0.10	100.00	4.00	0.40	5.69	3.35	5.93	5.29

pothesis testing procedures. The importance and relevance of the multiple hypothesis testing problem is well-recognized, and it is one of the current research challenges in Statistics because of the need to deal with the proliferation of high-dimensional "large M, small n" data sets, which are being created or generated due to advances in high-throughput technology. These technological advances, which is fueled by developments in computer technology and miniaturization, is embodied and spearheaded by, but not limited to, microarray technology. Appropriate statistical methods for such data sets need to take into account the undesirable impact of multiplicity.

Almost a century ago, Neyman and Pearson demonstrated the need to take into account the power function, and the alternative hypothesis configuration, when one is seeking an optimal test procedure in the one-pair hypothesis setting. Their work led to a divorce from the then-existing significance or p-value approach. Currently, many multiple hypothesis testing procedures, epitomized by the Sidak and Bonferonni procedures for control of the FWER, and by the famous Benjamini-Hochberg (BH) procedure for control of the FDR, are based on the p-values of the individual tests and do not seem to consider possible differences in the powers of the individual tests.

In this paper we examined the question of whether differences in the power characteristics of the individual tests in a multiple hypothesis testing situation could be exploited to improve on existing procedures for FWER and FDR control. This was done in a general decision-theoretic framework to allow for results that are applicable even with more complicated data types and structures. It was found that an optimal procedure within a certain class of procedures exists for weak FWER control, and this procedure exploits differences in the power characteristics of the individual tests. In particular, this procedure is better than the Sidak procedure, though the latter is a special case when all the power functions are identical. By utilizing this optimal FWER-controlling procedure as an anchor, a generalization of the BH procedure was constructed. This was shown to control the FDR and by the nature of its construction this was expected to have a smaller MDR than the BH procedure. A modest simulation study reveals that this new procedure has a smaller MDR than the BH procedure, at least for the simulation model that was considered. When the power functions of the individual tests are identical and with continuous data, this new procedure reduces to the BH procedure.

However, we cannot claim that the procedures developed in this paper are the best among all possible multiple decision functions, either for FWER or FDR control. The reason for this is that in constructing the new procedures, the starting point of the construction is a class of decision functions in which the component decision function for each gene *only* depends on data for that gene. So, even though the generalized BH multiple decision function is eventually adaptive, with each component depending on the whole data, it is not yet clear whether it will possess any global optimality property relative to the class of all multiple decision functions controlling the FDR. This appears to be a non-trivial problem since there will then be a need to characterize the latter class of multiple decision functions.

A natural layer to add in the decision-theoretic formulation of the problem is a Bayesian layer where a prior measure is specified on the unknown probability measure Q or, alternatively, on $\theta(Q)$. There is a possibility that through this Bayesian approach, one may be able to obtain a characterization of the class of optimal procedures controlling a certain type I error rate, or when the two types of error rates are combined in some way, for example, a weighted linear combination. This approach, however, requires a deeper investigation.

Acknowledgements. Many ideas and results in this paper were obtained while the first author was a Faculty Fellow at the Statistical and Applied Mathematical Sciences Institute (SAMSI) during his sabbatical leave from the University of South Carolina in Fall 2008. He is grateful to SAMSI Director Dr. James Berger for facilitating his stay at the Institute. As such, this work was partially supported by the National Science Foundation (NSF) under Grant DMS-0635449 to SAMSI. However, any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Some of the main results on this paper were obtained while the first author was visiting the University of Oslo in Oslo, Norway and the Norwegian University of Science and Technology (NTNU) in Trondheim, Norway in Fall 2008. He wishes to thank Professor Odd Aalen and Professor Bo Lindqvist for facilitating his visit to these universities. The discussions with colleagues in these institutions, such as Professors Ornulf Borgan, Nils Hjort, Havard Rue, and Gunnar Taraldsen, aside from Professors Aalen and Lindqvist, were beneficial.

The first author also wishes to acknowledge the following grants which partially support his research: NSF Grant DMS 0805809, National Institutes of Health (NIH) Grant RR17698, and Environmental Protection Agency (EPA) Grant RD-83241902-0 to University of Arizona with subaward number Y481344 to the University of South Carolina.

Finally, the authors wish to acknowledge helpful discussions with Professor James Lynch, Dr. Alex McLain, Dr. Marge Peña, Professor Jayaram Sethuraman, Dr. Laura Taylor, and Ms. Wensong Wu. They also thank Professor Larry Wasserman for alerting them to the martingale proof in (31).

References.

[1] Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B* **57**, 1, 289–300. MRMR1325392 (96d:62143)

[2] Doob, J. L. (1953). *Stochastic processes*. John Wiley & Sons Inc., New York. MRMR0058896 (15,445b)

[3] Dudoit, S., Gilbert, H. N., and van der Laan, M. (2007). Resampling-based empirical bayes multiple testing procedures for controlling generalized tail probability and expected value error rates: Focus on the false discovery rate and simulation study. Tech. rep., University of California at Berkeley.

[4] Dudoit, S., Shaffer, J. P., and Boldrick, J. C. (2003). Multiple hypothesis testing in microarray experiments. *Statist. Sci.* **18**, 1, 71–103. MRMR1997066

[5] Dudoit, S. and van der Laan, M. J. (2008). *Multiple testing procedures with applications to genomics*. Springer Series in Statistics. Springer, New York. MRMR2373771

[6] Efron, B. (2004). Large-scale simultaneous hypothesis testing: the choice of a null hypothesis. *Journal of the American Statistical Association 99*, 96–104.

[7] Efron, B. (2007). Size, power and false discovery rates. *Ann. Statist.* **35**, 4, 1351–1377. MRMR2351089

[8] Efron, B. (2008). Microarrays, empirical Bayes and the two-groups model. *Statist. Sci.* **23**, 1, 1–22. MRMR2431866

[9] Efron, B., Tibshirani, R., Storey, J. D., and Tusher, V. (2001). Empirical Bayes analysis of a microarray experiment. *J. Amer. Statist. Assoc.* **96**, 456, 1151–1160. MRMR1946571

[10] Genovese, C. and Wasserman, L. (2002). Operating characteristic and extensions of the false discovery rate procedure. *Journal of the Royal Statistical Society, Series B 64*, 499–517.

[11] Genovese, C. R., Roeder, K., and Wasserman, L. (2006). False discovery control with p-value weighting. *Biometrika* **93**, 3, 509–524. MRMR2261439 (2008c:62018)

[12] Hoeffding, W. (1956). On the distribution of the number of successes in independent trials. *Ann. Math. Statist. 27*, 713–721. MRMR0080391 (18,240f)

[13] Ihaka, R. and Gentleman, R. (1996). R: a language for data analysis and graphics. *Journal of Computational and Graphical Statistics 5*, 299–314.

[14] Jin, J. and Cai, T. T. (2007). Estimating the null and the proportional of nonnull effects in large-scale multiple comparisons. **102**, 478, 495–506.

[15] Langaas, M., Lindqvist, B. H., and Ferkingstad, E. (2005). Estimating the proportion of true null hypotheses, with application to DNA microarray data. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **67**, 4, 555–572. MRMR2168204

[16] Lehmann, E. L. (1997). *Testing statistical hypotheses*, second ed. Springer Texts in Statistics. Springer-Verlag, New York. MRMR1481711 (98m:62004)

[17] Marshall, A. W. and Olkin, I. (1979). *Inequalities: theory of majorization and its applications*. Mathematics in Science and Engineering, Vol. **143**. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York. MRMR552278 (81b:00002)

[18] Neyman, J. and Pearson, E. (1933). On the problem of the most efficient tests of statistical hypotheses. *Phil. Trans. Roy. Soc. Ser. A 231*, 289–337.

[19] Rogosa, D. (2003). Accuracy of api index and school base report elements: 2003 academic performance index, california department of education. *Available at http://www.cde.cagov/ta/ac/ap/researchreports.asp*.

[20] Royden, H. (1968). *Real Analysis*, Second ed. McMillan Publishing Co., Inc., New York.

[21] Rubin, D., Dudoit, S., and van der Laan, M. (2006). A method to increase the power of multiple testing procedures through sample splitting. *Stat. Appl. Genet. Mol. Biol. 5*, Art. 19, 20 pp. (electronic). MRMR2240850

[22] Rubin, D., van der Laan, M., and Dudoit, S. (2005). Multiple testing procedures which are optimal at a simple alternative. Tech. rep., University of California at Berkeley.

[23] Schwartzman, A., Dougherty, R., and Taylor, J. (2005). Cross-subject comparison of principal diffusion direction maps. *Magnetic Resonance in Medicine 53*, 1423–1431.

[24] Schweder, T. and Spjotvoll, E. (1982). Plots of p-values to evaluate many tests simultaneously. *Biometrika* **69**, 3, 493–502. http://www.jstor.org/stable/2335984.

[25] Šidák, Z. (1967). Rectangular confidence regions for the means of multivariate normal distributions. *J. Amer. Statist. Assoc. 62*, 626–633. MRMR0216666 (35 #7495)

[26] Singh, D., Febbo, P., Ross, K., Jackson, D., Manola, J., Ladd, C., Tamayo, P., Renshaw, A., D'Amico, A., Richie, J., Lander, E., Loda, M., Kantoff, P., Golub, T., and Sellers, R. (2002). Gene expression correlates of clinical prostate cancer behavior. *Cancer Cell 1*, 302–309.

[27] Spjøtvoll, E. (1972). On the optimality of some multiple comparison procedures. *Ann. Math. Statist. 43*, 398–411. MRMR0301871 (46 #1026)

[28] Storey, J. (2002). A direct approach to false discovery rates. *Journal of the Royal Statistical Society, Series B 64*, 479 – 498.

[29] Storey, J. (2003). The positive false discovery rate: a bayesian interpretation and the q-value. *The Annals of Statistics 31*, 2012 – 2035.

[30] Storey, J. (2007). The optimal discovery procedure: a new approach to simultaneous significance testing. *Journal of the Royal Statistical Society, Series B 69*, 347–368.

[31] Storey, J. D., Taylor, J. E., and Siegmund, D. (2004). Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **66**, 1, 187–205. MRMR2035766 (2004k:62056)

[32] Sun, W. and Cai, T. (2007). Oracle and adaptive compound decision rules for false discovery rate control. *Journal of the American Statistical Association 102*, 901–912.

[33] Turnbull, B. (2006). Best proteomics data. *Available at www.stanford.edu/people/brit.turnbull/BESTproteomics.pdf*.

[34] Wasserman, L. and Roeder, K. (2006). Weighted hypothesis testing. Tech. rep., Carnegie-Mellon University.

Department of Statistics
University of South Carolina
Columbia, SC 29208
E-mail: pena@stat.sc.edu
URL: http://www.stat.sc.edu/pena

Department of Statistics
University of South Carolina
Columbia, SC 29208
E-mail: habigerj@mailbox.sc.edu