Published in final edited form as:

Metrika. 2015 July 1; 78(5): 563–595. doi:10.1007/s00184-014-0516-6.

Classes of Multiple Decision Functions Strongly Controlling FWER and FDR

Edsel A. Peña,

Department of Statistics, University of South Carolina, Columbia, SC 29208 USA, Tel.: 803-576-5813, Fax: 803-777-4048, pena@stat.sc.edu

Joshua D. Habiger, and

Department of Statistics, Oklahoma State University, Stillwater, OK 74078, USA, jhabige@okstate.edu

Wensong Wu

Department of Mathematics and Statistics, Florida International University, Miami, FL 33199, USA, wenswu@fiu.edu

Abstract

Two general classes of multiple decision functions, where each member of the first class strongly controls the family-wise error rate (FWER), while each member of the second class strongly controls the false discovery rate (FDR), are described. These classes offer the possibility that optimal multiple decision functions with respect to a pre-specified Type II error criterion, such as the missed discovery rate (MDR), could be found which control the FWER or FDR Type I error rates. The gain in MDR of the associated FDR-controlling procedure relative to the well-known Benjamini-Hochberg (BH) procedure is demonstrated via a modest simulation study with gamma-distributed component data. Such multiple decision functions may have the potential of being utilized in multiple testing, specifically in the analysis of high-dimensional data sets.

Keywords

false discovery rate; family-wise error rate; missed discovery rate; multiple decision problem; multiple testing; strong control

1 Introduction and Motivation

Consider the problem of simultaneously testing M pairs of null and alternative hypotheses, (H_{m0}, H_{m1}) , $m = 1, 2, \dots, M$, where M is a known positive integer. Such multiple testing problems arise in many scientific areas; see, for instance, [1,2] for concrete examples. A discovery is said to have been made in the m th testing problem if H_{m0} is rejected. It is called a false discovery if a discovery is declared but in reality H_{m0} is true. On the other hand, a nondiscovery is made at the m th testing problem if H_{m0} is not rejected. It is called a missed discovery if a nondiscovery is declared when in reality H_{m0} is false. For the simultaneous testing problem, two commonly-used global Type I error rates are the family-wise error rate

(FWER), which is the probability of at least one false discovery, and the false discovery rate (FDR), which is the expectation of the ratio of the number of false discoveries over the number of discoveries.

A conventional testing paradigm employed in these situations is to decide on the collection of statistical tests for the M pairs of hypotheses, e.g., a t -test or a Mann-Whitney-Wilcoxon test for each pair, obtain the p -value for each test, and then use the resulting M p -values in the FWER-controlling sequential Šidák procedure, provided an independence condition is satisfied, or the FDR-controlling procedure in [3]. In this conventional approach, there appears to be no leeway in the choice of the multiple testing procedure the moment the individual test procedures have been chosen. Furthermore, since p -values are probabilities computed under distributions specified by the null hypotheses, it is not apparent whether these p -value-based procedures are actually taking into account the probabilities under the alternative hypotheses distributions. If they do not, then it goes against the Neyman-Pearson dictum that in the search for optimal test procedures, it is germane to consider probabilities under both the null and the alternative hypotheses.

This paper is primarily motivated by these issues. In particular, we pose the following question: If we are given the M test procedures for each of the M pairs of hypotheses, could we obtain classes of multiple testing procedures whose elements either control the FWER or the FDR? If this has an affirmative answer, then a multiple testing procedure within these classes which is optimal with respect to some Type II error rate criterion may be found. In turn, we may then be able to choose the starting collection of test functions that will provide the best multiple testing procedure. This is the spirit of this paper. We will in fact demonstrate that, under certain conditions, when given a collection of test functions for the M pairs of hypotheses, we can generate classes of multiple testing procedures controlling the FWER or the FDR. These results will have important implications in the search for optimal multiple testing procedures that control either of these Type I error rates. The main results in this paper were motivated by those in [4], which did not deal with classes of multiple testing procedures, but instead simply focussed in developing improved FWER and FDR-controlling procedures from the Neyman-Pearson most powerful tests for each of the M pairs of hypotheses.

We shall investigate these issues in a somewhat general abstract framework. We have opted for the additional abstraction in our mathematical framework in order to make concise and elegant the derivations and also to cover more general data structures that arise or may arise in such multiple testing problems. Such data structures are typically high-dimensional and of more complicated forms (e.g., data for the m th testing problem need not be a multi-sample data set but it could be a spatio-temporal, an image, or a shape data), such as those arising in neuroscience, genomics, proteomics, etc. However, in order to make the general results more accessible to the reader, we will demonstrate them in concrete and conventional settings. In line with this, we first discuss a concrete situation in Section 2 as a way of introducing issues of interest and to describe in this specific situation the major results of the paper.

There are certainly papers that have tried to improve on the p -value based approaches. One approach was to use weighted p -values such as in [5] which provides a FWER controlling procedure and in [6] which gives a procedure that controls the FDR. A variety of p -value weighting schemes have been proposed, usually relying on a posited model for the p -values and/or whether control of the FWER or FDR is desired; see, for instance, [7] for a review. Though our proposed testing procedures may be viewed as weighted p -value based, we mention at the outset that our approach is intrinsically different from the weighted p -value approach as expounded in [6] in the manner in which the p -value statistics are weighted. In [6], it is stated in their section 2 that:

Whatever information one uses to construct p -value weights, the weight assignment remains a guess. This guess is to be made a priori, that is before seeing the p -values. For purposes of analysis, we model the weights as random variables that are related to the underlying truth or falsehood of each null hypothesis.

In contrast, in our approach, the p -value weights arise from the alternative hypotheses of most interest. This is akin to determining an appropriate sample size to achieve a desired power for an alternative hypothesis of most interest. Our weights thus take into major consideration the powers of the individual tests at specified alternative hypotheses, hence are *not* based on a priori assessments of the truth or falsity of each of the null hypotheses.

There are other papers that provide general FWER- and FDR-controlling procedures and which aim for some optimality properties. The paper [8] provides two sufficient conditions for FDR control, while [9] concerns aspects of optimality. The papers [10,11] present sequential rejection procedures which control FWER. Recent papers dealing with multiple decision functions (MDFs) with certain optimality properties are [12,5,13,6,14–19]. On the other hand, papers proposing MDFs with a Bayes or empirical Bayes flavor, are [20,21,1,22]; and more recently, [23,24].

2 Essence of General Results via Concrete Situation

Prior to embarking on our abstract development we first provide a concrete situation to illuminate the issues and to offer a glimpse of the major results in this paper. Consider the problem of simultaneously testing M pairs of hypotheses $H_{m0} : \theta_m \leq \theta_{m0}$ versus $H_{m1} : \theta_m > \theta_{m0}$ ($m \in \mathcal{M} \equiv \{1, 2, \dots, M\}$), where θ_{m0} 's are known positive constants. The tests are to be based on data $\mathbf{T} = (T_m, m \in \mathcal{M})$, where T_m has a gamma distribution with shape parameter n_m , a known positive integer, and scale parameter θ_m , so that the marginal density function of T_m is

$$f_{T_m}(t|\theta_m) = \frac{1}{\theta_m^{n_m} \Gamma(n_m)} t^{n_m-1} \exp\{-t/\theta_m\} I\{t>0\},$$

where $I\{\cdot\}$ is the indicator function and $\Gamma(\cdot)$ is the gamma function. The Sufficiency Principle allows us to simply focus on the T_m 's since if $T_{m1}, T_{m2}, \dots, T_{mn_m}$ are the observable independent and identically distributed (IID) exponential random variables with scale parameter θ_m , then $T_m = \sum_{j=1}^{n_m} T_{mj}$ is sufficient for θ_m and has a gamma distribution

with parameter vector (n_m, θ_m) . Henceforth, we shall denote by $\mathcal{G}(\cdot; n, \theta)$ and $\mathcal{G}^{-1}(\cdot; n, \theta)$ the distribution and quantile functions, respectively, of a gamma random variable with parameter vector (n, θ) . We may also assume that $(T_m, m \in \mathcal{M}_0)$ is an independent collection of random variables, where $\mathcal{M}_0 = \{m \in \mathcal{M} : H_{m0} \text{ is true}\}$. On the other hand, $(T_m, m \in \mathcal{M}_1)$, where $\mathcal{M}_1 = \mathcal{M} \setminus \mathcal{M}_0$, need not be an independent collection of random variables. It is well-known that, marginally, the α_m -size uniformly most powerful (UMP) test for $H_{m0} : \theta_m \leq \theta_{m0}$ versus $H_{m1} : \theta_m > \theta_{m0}$ based on T_m is given by

$$\delta_m(T_m; \alpha_m) = I\{T_m \geq \mathcal{G}^{-1}(1 - \alpha_m; n_m, \theta_{m0})\}.$$

Recall that a (non-randomized) test function δ is a $\{0, 1\}$ -valued statistic such that $\delta = 0(1)$ means that H_{m0} is not rejected (rejected). The collection $\delta = (\delta_m, m \in \mathcal{M})$, which is a mapping from the sample space $\mathcal{R}_+^M$ of the $(T_m, m \in \mathcal{M})$ into $\{0, 1\}^M$, is a specific case of a multiple decision or test function.

Since we are now performing M simultaneous tests, *global* measures of Type I error are required. The first such global measure of Type I error is the family-wise error rate (FWER). For the multiple testing function $\delta \equiv \{\delta_m(\cdot; \alpha_m), m \in \mathcal{M}\}$, its FWER is

$$FWER(\delta) = \mathbf{P} \left\{ \sum_{m \in \mathcal{M}_0} \delta_m(T_m; \alpha_m) \geq 1 \right\}$$

where $\mathbf{P}$ is the true probability measure governing the $(T_m, m \in \mathcal{M})$. Another global measure of Type I error in the simultaneous testing setting is the false discovery rate (FDR), which for the multiple testing function δ is defined via

$$FDR(\delta) = E_{\mathbf{P}} \left\{ \frac{\sum_{m \in \mathcal{M}_0} \delta_m(T_m; \alpha_m)}{\max\{\sum_{m \in \mathcal{M}} \delta_m(T_m; \alpha_m), 1\}} \right\}.$$

We emphasize that the true probability measure $\mathbf{P}$ in the probabilistic evaluation in the FWER or the expectation evaluation in the FDR is not known. Furthermore, note that the set of indices $\mathcal{M}_0$ also depends on this unknown $\mathbf{P}$. Without knowing this $\mathbf{P}$, if one is able to establish that either the FWER or the FDR is no more than a fixed specified level $q \in [0, 1]$ by proper choice of the individual sizes α_m 's, or perhaps by proper construction of a multiple testing procedure δ which may be of much more complicated form than the $(\delta_m(\cdot; \alpha_m), m \in \mathcal{M})$, e.g., each component δ_m may depend on all the T_m 's or it could have a sequential flavor, then *strong* control of either of these global Type I error measures is achieved.

There are existing multiple test functions that control either the FWER or the FDR which are anchored on the p -value statistics obtained from the individual test functions δ_m 's. For controlling the FWER we may, for instance, utilize the sequential Šidák procedure (see, also, [10]); while for FDR control a popular multiple test function is the Benjamini and Hochberg [3] (BH) procedure. In the context of the concrete situation described in this

section, the main essence and contribution of this paper is the demonstration that, starting from the individual UMP tests δ_m 's, one could construct a *class* of multiple test functions whose members control the FWER and with this class including as a special case the Šidák procedure; as well as a *class* of multiple test functions whose members control the FDR and with the BH procedure being in fact a member of this class. We now describe *specific* subsets of these classes of multiple test functions. Let

$$\Gamma = \left\{ (a_1, a_2, \dots, a_M) : a_m \in [0, 1]; \sum_{m=1}^M a_m = 1 \right\}.$$

Let $\gamma = (\gamma_m, m \in \mathcal{M}) \in \Gamma$. For each $m \in \mathcal{M}$, define the function $A_m : [0, 1] \rightarrow [0, 1]$ via

$$A_m(\alpha; \gamma_m) = 1 - (1 - \alpha)^{\gamma_m}. \quad (1)$$

Suppose that it is desired to control the FWER at level $q \in (0, 1)$. For this purpose, define the statistic

$$\alpha^\dagger(\mathbf{T}; q, \gamma) = \inf \left\{ \alpha \in [0, 1] : \prod_{m \in \mathcal{M}} [1 - A_m(\alpha; \gamma_m)]^{1 - \delta_m[T_m; A_m(\alpha; \gamma_m)]} < 1 - q \right\},$$

where $\delta_m(t_m; \alpha-) \equiv \lim_{\beta \uparrow \alpha} \delta_m(t_m; \beta)$. Then, Theorem 1 on page 16 guarantees us that the multiple test function given by

$$\delta^\dagger(\mathbf{T}; q, \gamma) = (\delta_m(T_m; A_m(\alpha^\dagger(\mathbf{T}; q, \gamma); \gamma_m)), m \in \mathcal{M}) \quad (2)$$

in fact controls the FWER at q for *any* fixed $\gamma \in \Gamma$.

On the other hand, suppose that it is desired to control the FDR at level q . If we define the statistic

$$\alpha^*(\mathbf{T}; q, \gamma) = \sup \left\{ \alpha \in [0, 1] : \sum_{m \in \mathcal{M}} A_m(\alpha; \gamma_m) \leq q \sum_{m \in \mathcal{M}} \delta_m[T_m; A_m(\alpha; \gamma_m)] \right\},$$

then Theorem 2 on page 16 assures us that the multiple test function

$$\delta^*(\mathbf{T}; q, \gamma) = (\delta_m(T_m; A_m(\alpha^*(\mathbf{T}; q, \gamma); \gamma_m)), m \in \mathcal{M}) \quad (3)$$

controls the FDR at level q for $\gamma \in \Gamma$ satisfying additional conditions to be stated later.

Observe that both $\delta^\dagger(\mathbf{T}; q, \gamma)$ and $\delta^*(\mathbf{T}; q, \gamma)$ are so-called compound multiple test functions since their m th component functions utilize the whole data vector $\mathbf{T}$, not just T_m .

If we take $\gamma_m = 1/M$ for each $m \in \mathcal{M}$, then the multiple test function $\delta^\dagger(\mathbf{T}; q, \gamma = \mathbf{1}/M)$ is the sequential Šidák procedure; while the multiple test function $\delta^*(\mathbf{T}; q, \gamma = \mathbf{1}/M)$ is the BH procedure. But since by varying the vector $\gamma = (\gamma_m, m \in \mathcal{M})$ we obtain many FWER- or FDR-controlling multiple test functions, there is now the possibility that by considering the

different power characteristics of the individual test functions δ_m 's, possibly owing to differences in the values of θ_{m0} 's, n_m 's and/or the effect sizes that we would like to detect, we may now be able to choose appropriately the γ_m 's to improve the global power of the resulting multiple test functions $\delta^\dagger(\mathbf{T}; q, \gamma)$ or $\delta^*(\mathbf{T}; q, \gamma)$. The FWER-control or FDR-control for such a choice of γ_m 's are then guaranteed to be satisfied by virtue of the results concerning the class of multiple test functions. In a nutshell this concrete setting provides a glimpse of the essence of this paper. This concrete setting is also used in the simulation study presented in subsection 7.2.

Since the problem of multiple testing is truly a general problem with potentially complicated data structures, we have opted to treat this more abstractly. Undeniably, it makes the paper denser and harder to comprehend; however, through the abstraction, the results obtained and the proofs presented are more general in scope.

3 Mathematical Underpinnings

3.1 Decision-Theoretic Elements

Let $(\mathcal{X}, \mathcal{F}, \mathcal{P})$ be a statistical model, so $(\mathcal{X}, \mathcal{F})$ is a measurable space and $\mathcal{P}$ is a collection of probability measures on $(\mathcal{X}, \mathcal{F})$. Though not needed in the abstract development, we may adopt for concreteness the interpretation that $\mathcal{X}$ is the range space of an observable random entity X arising from an experiment. We shall denote by $\mathbf{P} \in \mathcal{P}$ the true underlying, but unknown, probability measure on $(\mathcal{X}, \mathcal{F})$. All probability statements and expectations will therefore be with respect to $\mathbf{P}$, while a statistical hypothesis will be a proposition regarding $\mathbf{P}$. In the sequel, for a space $\mathcal{S}$, $\sigma(\mathcal{S})$ will denote an appropriate sigma-field of subsets of $\mathcal{S}$. In decision problems with action space $\mathcal{A} = \{0, 1\}$, such as in hypothesis testing, a nonrandomized decision function is a $\delta : (\mathcal{X}, \mathcal{F}) \rightarrow (\mathcal{A}, \sigma(\mathcal{A}))$. In hypothesis testing, given $X = x \in \mathcal{X}$, a decision $\delta(x) = 0$ corresponds to deciding in favor of the null hypothesis (H_0), whereas $\delta(x) = 1$, a so-called *discovery*, corresponds to deciding in favor of the alternative hypothesis (H_1).

It suffices to restrict ourselves to nonrandomized decision functions since, through the use of an auxiliary randomizer, usually a standard uniform variable U independent of X , we may always convert a randomized decision function $\delta^* : (\mathcal{X}, \mathcal{F}) \rightarrow ([0, 1], \sigma[0, 1])$ into a nonrandomized decision function $\delta : (\mathcal{X} \times [0, 1], \mathcal{F} \otimes \sigma[0, 1]) \rightarrow (\mathcal{A}, \sigma(\mathcal{A}))$ via $\delta(x, u) = I\{\delta^*(x) \leq u\}$. Thus, in our general formulation, the sample space $\mathcal{X}$ may actually represent a product space between a data space and $[0, 1]$. This framework is appropriate, for instance, when dealing with discrete data or when using nonparametric rank-based decision functions. For more discussions on this matter, see [4, 25].

Decision or test functions depend on a size parameter $\alpha \in [0, 1]$ as in the concrete situation in Section 2. To further demonstrate this idea, consider the problem of testing the null hypothesis $H_0 : \mu = 0$ versus the alternative hypothesis $H_1 : \mu > 0$ based on a random observable $X \sim N(\mu, 1)$, where $N(a, b^2)$ is a normal distribution with mean a and variance b^2 . Note that it suffices to consider this one-dimensional X by virtue of a Sufficiency reduction. The size- α test $\delta : \mathcal{X} \equiv \mathbb{R} \rightarrow \{0, 1\}$ has $\delta(x; \alpha) = I\{|x| > \Phi^{-1}(1 - \alpha/2)\}$, where $\Phi^{-1}(\cdot)$ is the quantile function of the standard normal distribution. Henceforth, to simplify

our notation, we adopt a functional notation where $\delta(\alpha)$ represents the statistic defined on $\mathcal{X}$ according to $x \mapsto \delta(x; \alpha)$. When we view this as a stochastic process in α , we obtain the notion of a (nonrandomized) decision process as introduced in [4], which is a stochastic process $\delta = \{\delta(\alpha) : \alpha \in [0, 1]\}$ where, $\forall \alpha \in [0, 1]$, $\delta(\alpha)$ is a decision function, and such that the following conditions are satisfied.

- (D1) $\delta(0) = 0$ and $\delta(1) = 1$ a.e.- $\mathbf{P}$.
- (D2) The sample paths $\alpha \mapsto \delta(\alpha)$ are, a.e.- $\mathbf{P}$, $\{0, 1\}$ -valued step-functions which are nondecreasing and right-continuous.

3.2 Multiple Decision Functions

Let $\mathcal{M}$ be a known finite set with $|\mathcal{M}| = M$. An $\mathcal{M}$ -indexed multiple decision problem is one whose action space is $\mathcal{A}^M$. In the context of multiple hypotheses testing, for each $m \in \mathcal{M}$, there is a pair of hypotheses H_{m0} and H_{m1} . Of interest is to simultaneously decide between H_{m0} and H_{m1} for each $m \in \mathcal{M}$. A multiple decision function (MDF) is a $\delta = (\delta_m : m \in \mathcal{M})$, where δ_m is a decision function. Thus, $\delta : (\mathcal{X}, \mathcal{F}) \rightarrow (\mathcal{A}^M, \sigma(\mathcal{A}^M))$. A multiple decision process (MDP) is a $\delta = (\delta_m : m \in \mathcal{M})$, where $\delta_m = \{\delta_m(\alpha) : \alpha \in [0, 1]\}$ is a decision process.

Let $\mathcal{M}_0 \equiv \mathcal{M}_0(\mathbf{P})$ and $\mathcal{M}_1 \equiv \mathcal{M}_1(\mathbf{P})$ be subsets of $\mathcal{M}$ such that

$$\mathcal{M} = \mathcal{M}_0(\mathbf{P}) \cup \mathcal{M}_1(\mathbf{P}) \text{ and } \mathcal{M}_0(\mathbf{P}) \cap \mathcal{M}_1(\mathbf{P}) = \emptyset.$$

We shall assume that the following condition holds.

- (D3) $\{\delta_m : m \in \mathcal{M}_0(\mathbf{P})\}$ and $\{\delta_m : m \in \mathcal{M}_1(\mathbf{P})\}$ are independent of each other and the elements of $\{\delta_m : m \in \mathcal{M}_0(\mathbf{P})\}$ are independent.

In multiple hypotheses testing, H_{m0} is true under $\mathbf{P}$ if and only if $m \in \mathcal{M}_0(\mathbf{P})$. Observe that the elements of $\{\delta_m : m \in \mathcal{M}_1(\mathbf{P})\}$ need not be independent. In many cases, such as in the illustration in Section 2, the δ_m test process may just be a function of X_m which is specific for the m th decision problem. Condition (D3) will then be satisfied if $\{X_m, m \in \mathcal{M}_0(\mathbf{P})\}$ is an independent collection. However, there are cases where the M test processes may be using the same data X , and in such a case we require that the independence condition in (D3) should hold.

In addition, we shall also assume the condition that

- (D4) $\forall m \in \mathcal{M}_0(\mathbf{P}), \forall \alpha \in [0, 1]: E_{\mathbf{P}} \{\delta_m(\alpha)\} = \alpha$. The collection of all $\mathcal{M}$ -indexed multiple decision processes satisfying conditions (D1)–(D4) will be denoted by $\mathfrak{D}$. We remark that the requirement of equality in (D4), given by $E_{\mathbf{P}} \{\delta_m(\alpha)\} = \alpha$, will be fulfilled in many situations since an auxiliary randomizer is incorporated in our framework. However, there may still be situations when dealing with non-regular families of distributions (e.g., uniform family of distributions) where this condition may not be satisfied. The latter will manifest itself when the decision function achieves power one but with its size not yet equal to one.

3.3 Multiple Decision Size Functions

Let $\mathbf{A} = (A_m : m \in \mathcal{M})$ be an $\mathcal{M}$ -indexed collection of measurable functions with $A_m : ([0, 1], \sigma[0, 1]) \rightarrow ([0, 1], \sigma[0, 1])$. We shall say that $\mathbf{A}$ is a multiple decision size function if it satisfies the following three conditions:

$$(A1) \quad \forall m \in \mathcal{M}: A_m(0) = 0 \text{ and } A_m(1) = 1.$$

$$(A2) \quad \forall m \in \mathcal{M}: \alpha \mapsto A_m(\alpha) \text{ is strictly increasing and continuous.}$$

$$(A3) \quad \forall \alpha \in [0, 1] : \prod_{m \in \mathcal{M}} [1 - A_m(\alpha)] = 1 - \alpha.$$

The idea behind the introduction of these size functions is that they will serve as ‘size-pickers’ for each of the individual test functions when a global Type I error level α is specified. Thus, given an α , the multiple decision function that will be of interest is $(\delta_m[A_m(\alpha)], m \in \mathcal{M})$. Conditions (A1) and (A2) are intuitive requirements for size-pickers. On the other hand, Condition (A3) guarantees that the multiple decision size function $(\delta_m[A_m(\alpha)], m \in \mathcal{M})$ achieves a *weak* FWER of no more than α . The weak FWER is the FWER under a $\mathbf{P}$ such that all the H_{m0} ’s are true. We denote the collection of all $\mathcal{M}$ -indexed multiple decision size functions by $\mathfrak{S}$.

When we deal with FDR-controlling multiple decision functions, we need an additional condition that controls the interplay among the multiple decision process, the multiple size function $\mathbf{A}$, and the underlying probability measure $\mathbf{P}$. This condition is as follows:

$$(C) \quad \text{The multiple decision process } \mathbf{d} = \{d_m : m \in \mathcal{M}\}, \text{ multiple size function } \mathbf{A} = \{A_m : m \in \mathcal{M}\}, \text{ and underlying probability measure } \mathbf{P} \text{ satisfy, whenever } |\mathcal{M}_0(\mathbf{P})| < M,$$

$$\sup_{\alpha \in [\alpha_{(1)}, 1]} \left[\frac{|\mathcal{M}_0(\mathbf{P})| \max_{m \in \mathcal{M}_0(\mathbf{P})} A_m(\alpha)}{\sum_{m \in \mathcal{M}(\mathbf{P})} A_m(\alpha)} \right] \leq 1, \text{ a.s. } [\mathbf{P}],$$

with

$$\alpha_{(1)} \equiv \alpha_{(1)}(\mathbf{d}, \mathbf{A}) = \inf \left\{ \alpha \in [0, 1] : \sum_{m \in \mathcal{M}} \delta_m[A_m(\alpha)] \geq 1 \right\}.$$

Note that $\alpha_{(1)}$ is a random variable, and it is the minimum of the so-called generalized p -value statistics which are formally defined in (12). This statistic $\alpha_{(1)}$ could be interpreted as the smallest global error rate that leads to a rejection of at least one H_{m0} ’s when using the multiple decision process $\mathbf{d}$ and multiple size function $\mathbf{A}$. Verifying the condition in (C) is not a trivial matter because of the randomness of $\alpha_{(1)}$ whose distribution depends on the underlying unknown probability measure $\mathbf{P}$. A stronger but perhaps easier to verify condition, which implies the condition in (C), is that

$$\sup_{\alpha \in [0, 1]} \left[\frac{|\mathcal{M}_0(\mathbf{P})| \max_{m \in \mathcal{M}_0(\mathbf{P})} A_m(\alpha)}{\sum_{m \in \mathcal{M}(\mathbf{P})} A_m(\alpha)} \right] \leq 1. \quad (4)$$

As mentioned above, condition (C) depends on the true probability measure $\mathbf{P}$. Since $\mathbf{P}$ is not known, we do not know $\mathcal{M}_0(\mathbf{P})$. However, we may have an idea, possibly justified by *sparsity of signals* considerations, that the ratio M_0/M , where $M_0 = |\mathcal{M}_0(\mathbf{P})|$, is *no more than* some value B with $0 < B < 1$. Condition (C) is then implied by

$$\forall \alpha \in [\alpha_{(1)}, 1]: \left[\frac{\max_{m \in \mathcal{M}} A_m(\alpha)}{\bar{A}(\alpha)} \right] \leq \frac{1}{B}, \text{ a.s.}[\mathbf{P}], \quad (5)$$

where $\bar{A}(\alpha) = \max_{m \in \mathcal{M}} A_m(\alpha)/M$. Visually, this condition implies that each of the size functions must be inside the upper envelope $U(\alpha)$ determined by

$$\alpha \in [0, 1] \mapsto U(\alpha) \equiv \frac{\bar{A}(\alpha)}{B}$$

for $\alpha \in [\alpha_{(1)}, 1]$. In essence, none of the size functions should dominate the other size functions, which is akin to Noether's condition when dealing with the asymptotics of rank-based statistics.

A particular element of $\mathfrak{S}$ is the Šidák multiple decision size function (cf., [26])

$\mathbf{A}^S = (A_m^S : m \in \mathcal{M})$ with

$$A_m^S(\alpha) = 1 - (1 - \alpha)^{1/M}, \alpha \in [0, 1], m \in \mathcal{M}. \quad (6)$$

This Šidák size function will play a central role in the proof of Theorem 2 which deals with multiple decision functions controlling the false discovery rate. Note, however, that the

Bonferroni size function $\mathbf{A}^B = (A_m^B : m \in \mathcal{M})$ with

$$A_m^B(\alpha) = \alpha/M, \alpha \in [0, 1], m \in \mathcal{M}, \quad (7)$$

does not belong to $\mathfrak{S}$ since $A_m(1) < 1$, hence condition (A1) is not satisfied. The fact that this Bonferroni size function does not even satisfy condition (A1) is testament to its being so conservative to the extent that it could not be considered among multiple decision size functions that could potentially lead to optimal multiple decision functions. For, if in the extreme case, we set FWER equal to $\alpha = 1$, then any multiple decision function whose components are all identically equal to one will satisfy such an FWER control, and we could therefore set $A_m(1) = 1$ for all $m \in \mathcal{M}$.

The Šidák multiple decision size function does not depend on m . We provide examples of multiple size functions that depend on m . The first one is that considered in Section 2 where we have, for $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_M) \in \Gamma$,

$$A_m(\alpha; \gamma_m) = 1 - (1 - \alpha)^{\gamma_m}, m \in \mathcal{M}.$$

We remark that the γ_m 's may also depend on α , as will be the case in a later concrete example. The collection $\{A_m(\alpha; \gamma_m) : m \in \mathcal{M}\}$ clearly satisfies conditions (A1) and (A2),

and if $\max_{m \in \mathcal{M}} \gamma_m = 1$, then condition (A3) is also satisfied. Equality in condition (A3) is actually achieved by having $\max_{m \in \mathcal{M}} \gamma_m = 1$. Condition (C) in this special case becomes:

$$\forall \alpha \in [\alpha_{(1)}, 1]: M_0 \max_{m \in \mathcal{M}_0} \{1 - (1 - \alpha)^{\gamma_m}\} \leq \sum_{m \in \mathcal{M}} \{1 - (1 - \alpha)^{\gamma_m}\}, \text{ a.s. } [\mathbf{P}].$$

Observe that if all H_{m0} 's are false, so that $\mathcal{M}_0 = \emptyset$, then the condition above is automatically satisfied; while if all the H_{m0} 's are true, so that $\mathcal{M}_0 = \mathcal{M}$, then we are forced to have $\gamma_m = 1/M$, $m = 1, 2, \dots, M$, which leads to the Šidák multiple size functions. In practice, we will have M_0/M bounded above by some number $B \in (0, 1)$, and in this case a sufficient condition for the above condition to hold is

$$\forall \alpha \in [\alpha_{(1)}, 1]: \frac{1 - (1 - \alpha)^{\gamma_{(M)}}}{1 - \frac{1}{M} \sum_{m=1}^M (1 - \alpha)^{\gamma_m}} \leq \frac{1}{B}, \text{ a.s. } [\mathbf{P}].$$

If the γ_m 's do *not* depend on α , then for this class of multiple size functions the sufficient condition in (4), via L'Hôpital's Rule and focusing on the limit as $\alpha \downarrow 0$, is implied by the

condition that $\frac{\gamma_{(M)}}{\bar{\gamma}} \leq \frac{1}{B}$, where $\bar{\gamma} = \frac{1}{M} \sum_{i=1}^M \gamma_m$. To see this, define the function g on $[0, 1]$ by

$$g(\alpha) \equiv g(\alpha; \gamma) = \frac{1 - (1 - \alpha)^{\gamma_{(M)}}}{\sum_{m=1}^M [1 - (1 - \alpha)^{\gamma_{(m)}}]}.$$

The assertion above follows if we could show that $g(\alpha)$ is a decreasing¹ function of α . This is the same as showing that

$$Mg(\alpha) = \frac{\sum_{m=1}^M [1 - (1 - \alpha)^{\gamma_{(M)}}]}{\sum_{m=1}^M [1 - (1 - \alpha)^{\gamma_{(m)}}]}$$

is decreasing in α . But this is immediate if we could show that, for every $m = 1, 2, \dots, M$,

$$\alpha \mapsto h(\alpha) \equiv \frac{1 - (1 - \alpha)^{\gamma_{(M)}}}{1 - (1 - \alpha)^{\gamma_{(m)}}}$$

is a decreasing function of $\alpha \in [0, 1]$. To show that this is decreasing in α , observe that for $0 < \gamma_{(m)} < \gamma_{(M)} = 1$, as α decreases from 1, $(1 - \alpha)^{\gamma_{(M)}}$ decreases at a faster rate than $(1 - \alpha)^{\gamma_{(m)}}$ and both are zero at $\alpha = 1$. Consequently, as α decreases from 1, $1 - (1 - \alpha)^{\gamma_{(M)}}$ decreases at a slower rate than $1 - (1 - \alpha)^{\gamma_{(m)}}$, and since both are in $[0, 1]$, then it follows that $h(\alpha)$ increases as α decreases from 1, that is, that $h(\alpha)$ is a decreasing function of α . Albeit a

¹Henceforth, decreasing will mean non-increasing, while increasing will mean non-decreasing

more formal way of showing that $h(\alpha)$ is decreasing in α is to show that its derivative is negative. We leave this alternative proof as a leisurely exercise for the curious reader.

The next example of a multiple decision size function is derived through an optimization of a global power of the multiple decision function. Let $X_m \sim N(\mu_m, 1)$ and consider testing $H_{m0} : \mu_m = 0$ versus $H_{m1} : \mu_m > 0$. The usual test function of size η_m for H_{m0} versus H_{m1} is

$$\delta_m(\eta_m) \equiv \delta_m(x_m; \eta_m) = I\{x_m > \Phi^{-1}(1 - \eta_m)\}.$$

Let $\{\gamma_m, m \in \mathcal{M}\}$ be a collection of known positive reals, which could be viewed as the *effect sizes* of interest. The power of δ_m at $\mu_m = \gamma_m$ is given by

$$\pi_m(\eta_m) = \Phi\left[\gamma_m - \Phi^{-1}(1 - \eta_m)\right].$$

The *weak* family wise error rate of the associated multiple decision function is

$$\text{FWER}(\eta_m, m \in \mathcal{M}) = \mathbf{P}_0 \left\{ \sum_{m \in \mathcal{M}} \delta_m(X_m; \eta_m) > 0 \right\} = 1 - \prod_{m \in \mathcal{M}} [1 - \eta_m],$$

where $\mathbf{P}_0$ is the probability measure associated with $\mu_m = 0$ for all $m \in \mathcal{M}$. Suppose we seek to control this weak FWER at an overall level $\alpha \in [0, 1]$, so that condition (A3) is necessarily satisfied, but at the same time maximize the overall power of the multiple decision at $\mu_m = \gamma_m, m = 1, 2, \dots, M$, given by

$$\text{POWER}(\eta_m, m \in \mathcal{M}) = \sum_{m \in \mathcal{M}} \pi_m(\eta_m).$$

Then, (cf., [4]) the multiple size function $\{A_m(\alpha), m \in \mathcal{M}\}$ should satisfy the two conditions: (i) for some $\lambda \in \mathfrak{R}$,

$$\forall m \in \mathcal{M} : (1 - A_m(\alpha)) \frac{\phi[\gamma_m - \Phi^{-1}(1 - A_m(\alpha))]}{\phi[\Phi^{-1}(1 - A_m(\alpha))]} = \lambda;$$

where $\phi(\cdot)$ is the standard normal density function, and (ii) $\sum_{m \in \mathcal{M}} \log[1 - A_m(\alpha)] = \log(1 - \alpha)$. If the effect sizes γ_m 's are not all identical, then these size functions are not identical, that is, they will depend on m . Note that some constraints will need to be imposed on the γ_m 's in order for the resulting size functions $\{A_m(\cdot) : m \in \mathcal{M}\}$ to satisfy condition (C).

Let us actually implement the optimality prescription above for a somewhat more restrictive family of multiple decision size functions given by the form

$$A_m(\alpha) \equiv A_m(\alpha; \kappa_m) = 1 - (1 - \alpha)^{\kappa_m}, m = 1, 2, \dots, M, \quad (8)$$

with $\kappa_m \in [0, 1]$. Utilizing the Gaussian setting of the preceding example, we therefore seek the optimal κ_m 's in order to control the weak FWER at level α and at the same time maximize the overall power. We will immediately impose the condition that $A_m(\alpha)$'s satisfy $\forall \alpha \in [0, 1]: 1 - \prod_{m=1}^M [1 - A_m(\alpha)] = \alpha$. Since $1 - A_m(\alpha) = (1 - \alpha)^{\kappa_m}$, $m = 1, 2, \dots, M$, then it follows that $\sum_{m=1}^M \kappa_m = 1$. Within this family of multiple size functions, the condition for optimality is that, for some $\lambda > 0$, we must have, for every $m = 1, 2, \dots, M$, that

$$(1 - \alpha)^{\kappa_m} \frac{\phi[\gamma_m - \Phi^{-1}((1 - \alpha)^{\kappa_m})]}{\phi[\Phi^{-1}((1 - \alpha)^{\kappa_m})]} = \lambda. \quad (9)$$

Letting $b_m \equiv b_m(\alpha) = b_m(\alpha; \kappa_m) = \Phi^{-1}((1 - \alpha)^{\kappa_m})$, $m = 1, 2, \dots, M$, the equation in (9) simplifies to

$$\log \Phi(b_m) + \gamma_m b_m = \log(\lambda) + \frac{1}{2} \gamma_m^2, m = 1, 2, \dots, M. \quad (10)$$

Observe that the condition $\sum_{m=1}^M \kappa_m = 1$ is equivalent to $\sum_{m=1}^M \log \Phi(b_m) = \log(1 - \alpha)$. Combining, we therefore obtain an expression for λ given by

$$\log(\lambda) = \frac{1}{M} \sum_{m=1}^M \gamma_m b_m - \frac{1}{2M} \sum_{m=1}^M \gamma_m^2 + \frac{1}{M} \log(1 - \alpha). \quad (11)$$

Substituting into (10), we get the set of equations determining the optimal b_m 's to be, for $m = 1, 2, \dots, M$,

$$\log \Phi(b_m) + \left[\gamma_m b_m - \frac{1}{M} \sum_{j=1}^M \gamma_j b_j \right] = \frac{1}{2} \left[\gamma_m^2 - \frac{1}{M} \sum_{j=1}^M \gamma_j^2 \right] + \frac{1}{M} \log(1 - \alpha).$$

We summarize these results into a proposition.

Proposition 1 For the family of multiple decision functions of form $A_m(\alpha) = 1 - (1 - \alpha)^{\kappa_m}$, $m = 1, 2, \dots, M$, the optimal κ_m 's for weak FWER control at α for the Gaussian model are given by

$$\kappa_m^*(\alpha) = \frac{\log \Phi[b_m^*(\alpha)]}{\log(1 - \alpha)}, m = 1, 2, \dots, M,$$

where the $b_m^*(\alpha)$'s solve in $\mathbf{b} = (b_1, b_2, \dots, b_M)^t$ the set of equations given by

$$\mathbf{Q}(\mathbf{b}) + [Dg(\gamma) - \frac{1}{M}(\mathbf{1}\gamma^t)]\mathbf{b} \left\{ \frac{1}{2} \left[Dg(\gamma^2) - \frac{1}{M}(\mathbf{1}(\gamma^2)^t) \right] + \frac{1}{M} \log(1 - \alpha) \right\} \mathbf{1},$$

where $Dg(\mathbf{a})$ represents the diagonal matrix with diagonal elements from vector $\mathbf{a}$, $\mathbf{1}$ is a $M \times 1$ vector of 1's, $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_M)^t$, $\gamma^2 = (\gamma_1^2, \gamma_2^2, \dots, \gamma_M^2)^t$, and

$$\mathbf{Q}(\mathbf{b}) = [\log \Phi(b_1), \log \Phi(b_2), \dots, \log \Phi(b_M)]^t.$$

Clearly, to compute the optimal b_m^* 's, and hence the κ_m^* 's and the optimal size functions, requires numerical methods. Furthermore, we point out that the optimal k_m^* 's depend on α . Thus, there arise some sort of dynamic adjusting of the optimal sizes to use for each of the decision functions in accordance with the weak FWER level α that is being used. We implemented the computation of the above functions using a Newton-Raphson procedure. For demonstration we took $M = 5$ and specified the effect sizes γ_m , $m = 1, 2, \dots, 5$, by taking the absolute values of five generated random variates from a standard normal distribution. The resulting effect sizes were $\gamma = (1.08, 1.43, 0.19, 1.73, 0.10)$. Figure 1 presents in the first plot frame the functions $b_m(\alpha)$, $m = 1, 2, \dots, 5$, the second plot frame provides the functions $A_m(\alpha)$, $m = 1, 2, \dots, 5$, and the third plot frame provides the functions $\kappa_m(\alpha)$, $m = 1, 2, \dots, M$, for α ranging from .001 to .999 with .001 increments. Observe that the size functions are intersecting which is due to the dependence of the optimal κ_m 's to the value of α , a manifestation of the dynamic adjusting that is alluded to above. We actually point out that the second plot frame provides the optimal size functions without restricting to the Šidák-type size functions. See Theorem 3 for the general set of equations for determining the optimal size functions and from which the results above is a special case.

We also recall the notion of generalized p -value statistics; see [4]. Given a $\mathbf{A} \in \mathfrak{D}$ and an $\mathbf{A} \in \mathfrak{S}$ for $m \in \mathcal{M}$ we define the random variable

$$\alpha_m \equiv \alpha_m(\mathbf{A}, \mathbf{A}) = \inf\{\alpha \in [0, 1] : \delta_m(A_m(\alpha)) = 1\}. \quad (12)$$

The collection $(\alpha_m(\mathbf{A}, \mathbf{A}) : m \in \mathcal{M})$ is called the vector of generalized p -value statistics associated with the pair $(\mathbf{A}, \mathbf{A})$. Observe that the usual p -value statistic associated with δ_m is $P_m = A_m(\alpha_m)$, hence the use of the adjective *generalized* for the α_m 's. We shall assume that these generalized p -values are a.e. $[\mathbf{P}]$ distinct. In the multiple testing literature, there is also the notion of *adjusted* p -value statistics; see, for instance, pages 32–34 of [27]. The adjusted p -value associated with the m th null hypothesis H_{m0} is defined to be the smallest global error rate (e.g., FWER, FDR) so that H_{m0} is still rejected. Examining the definition of the generalized p -value statistics, we note that $\alpha_m(\mathbf{A}, \mathbf{A})$ could also be viewed as the smallest global error rate α such that the m th test function $\delta[A_m(\alpha)]$ rejects H_{m0} . In some sense, therefore, the adjusted p -value statistics and the generalized p -value statistics are related in that both are global error rates leading to the rejection of null hypotheses. Perhaps, our notion of a generalized p -value statistic is more general than the notion of an adjusted p -value statistic since we are allowing more general multiple size functions as pickers of the specific test functions to use at each of the components of the multiple decision problem.

4 Main Theorems and Classes of MDFs

We shall present in this section the two main results that will enable the construction of the classes of multiple decision functions controlling FWER and FDR. Recall that $\mathbf{P}$ denotes the true, but unknown, underlying probability measure acting on the space $(\mathcal{X}, \mathcal{F})$.

Given $\mathbf{a} = \{a_m : m \in \mathcal{M}\} \in \mathfrak{D}$ and $\mathbf{A} = \{A_m : m \in \mathcal{M}\} \in \mathfrak{S}$ define the stochastic processes $\mathbf{S}_0 = \{S_0(\alpha) : \alpha \in [0, 1]\}$, $\mathbf{S} = \{S(\alpha) : \alpha \in [0, 1]\}$, and $\mathbf{F} = \{F(\alpha) : \alpha \in [0, 1]\}$ via

$$S_0(\alpha) \equiv S_0(\alpha; \mathbf{\Delta}, \mathbf{A}) = \sum_{m \in \mathcal{M}_0(\mathbf{P})} \delta_m(A_m(\alpha)); \quad (13)$$

$$S(\alpha) \equiv S(\alpha; \mathbf{\Delta}, \mathbf{A}) = \sum_{m \in \mathcal{M}} \delta_m(A_m(\alpha)); \quad (14)$$

$$F(\alpha) \equiv F(\alpha; \mathbf{\Delta}, \mathbf{A}) = \frac{S_0(\alpha)}{S(\alpha)} I\{S(\alpha) > 0\}, \quad (15)$$

with the convention that $0/0 = 0$. These quantities have the following interpretations. Given an $\alpha \in [0, 1]$, for each $m \in \mathcal{M}$, the decision function whose size is $A_m(\alpha)$ is chosen from $\mathcal{M}$, and the MDF $\delta(\alpha) \equiv (\delta_m[A_m(\alpha)] : m \in \mathcal{M})$ is employed in the decision-making. For this MDF $\delta(\alpha)$, $S_0(\alpha)$ is the number of false discoveries, $S(\alpha)$ is the number of discoveries, and $F(\alpha)$ is the proportion of false discoveries among all discoveries. Observe, however, that since $\mathbf{P}$ is unknown, both $\mathbf{S}_0$ and $\mathbf{F}$ are unobservable, whereas $\mathbf{S}$ is observable.

For $q \in [0, 1]$, let us also define the random variables

$$\alpha^\dagger(q) \equiv \alpha^\dagger(q; \mathbf{\Delta}, \mathbf{A}) = \inf \left\{ \alpha \in [0, 1] : \prod_{m \in \mathcal{M}} [1 - A_m(\alpha)]^{1 - \delta_m(A_m(\alpha))} < 1 - q \right\}; \quad (16)$$

and

$$\alpha^*(q) \equiv \alpha^*(q; \mathbf{\Delta}, \mathbf{A}) = \sup \left\{ \alpha \in [0, 1] : \sum_{m \in \mathcal{M}} A_m(\alpha) \leq q S(\alpha; \mathbf{\Delta}, \mathbf{A}) \right\}. \quad (17)$$

In essence, $\alpha^\dagger(q)$ is a *first* crossing-time random variable, whereas $\alpha^*(q)$ is a *last* crossing-time random variable. The forms of these two random variables were motivated and justified in Sections 6 and 7 in [4] for a specific multiple decision size function, but the justifications in that paper carry over to the more general setting considered here.

The two main results of this paper are contained in Theorem 1 and Theorem 2. We present these theorems, but defer their proofs to Section 5 after some discussion about their implications and potential usefulness.

Theorem 1 Under conditions (D1)–(D4) for $\mathfrak{D}$ and (A1)–(A3) for $\mathfrak{S}$

$$\forall \mathbf{\Delta} \in \mathfrak{D}, \forall \mathbf{A} \in \mathfrak{S}, \forall q \in [0, 1] : E_{\mathbf{P}} \{I\{S_0(\alpha^\dagger(q; \mathbf{\Delta}, \mathbf{A}); \mathbf{\Delta}, \mathbf{A}) \geq 1\}\} \leq q.$$

Observe that $E_{\mathbf{P}} \{I\{S_0(\alpha^\dagger(q; \mathbf{\Delta}, \mathbf{A}); \mathbf{\Delta}, \mathbf{A}) \geq 1\}\}$ is the FWER since it is the probability of committing at least one false discovery under $\mathbf{P}$. Thus, Theorem 1 shows that for any $q \in [0,$

1], any multiple decision process $\delta \in \mathfrak{D}$ and any multiple decision size function $\mathbf{A} \in \mathfrak{S}$ the MDF defined via

$$\delta^\dagger(q) \equiv \delta^\dagger(q; \Delta, \mathbf{A}) = (\delta_m[A_m(\alpha^\dagger(q; \Delta, \mathbf{A}))]: m \in \mathcal{M}), \quad (18)$$

strongly controls the FWER at q .

Theorem 2 Under conditions (D1)–(D4) for $\mathfrak{D}$ and (A1)–(A3) for $\mathfrak{S}$ then for every $\delta \in \mathfrak{D}$ and every $\mathbf{A} \in \mathfrak{S}$ satisfying condition (C),

$$\forall q \in [0, 1]: E_{\mathbf{P}}\{F(\alpha^*(q; \Delta, \mathbf{A}); \Delta, \mathbf{A})\} \leq q.$$

Note that $E_{\mathbf{P}}\{F(\alpha^*(q; \Delta, \mathbf{A}); \Delta, \mathbf{A})\}$ is the false discovery rate (FDR) as defined in the seminal paper of [3]. The implication of Theorem 2 is that if, for each $q \in [0, 1]$, and for any multiple decision process $\delta \in \mathfrak{D}$ and multiple decision size function $\mathbf{A} \in \mathfrak{S}$ satisfying (C), we define the MDF

$$\delta^*(q) \equiv \delta^*(q; \Delta, \mathbf{A}) = (\delta_m[A_m(\alpha^*(q; \Delta, \mathbf{A}))]: m \in \mathcal{M}), \quad (19)$$

then $\delta^*(q)$ is an MDF that (*strongly*) controls the FDR at q .

The importance of the preceding results is that each multiple decision process $\delta \in \mathfrak{D}$ may have an associated multiple decision size process $\mathbf{A} \equiv \mathbf{A}(\delta) \in \mathfrak{S}$ such that the resulting multiple decision functions $\delta^\dagger(q)$ or $\delta^*(q)$ possess some optimality property, for example, with respect to the missed discovery rate, a Type II error rate. To define this rate, let

$$M(\alpha) \equiv M(\alpha; \Delta, \mathbf{A}) = \frac{\sum_{m \in \mathcal{M}_1(\mathbf{P})} (1 - \delta_m(A_m(\alpha)))}{|\mathcal{M}_1(\mathbf{P})|} I\{|\mathcal{M}_1(\mathbf{P})| > 0\}. \quad (20)$$

The quantity $M(\alpha)$ has the interpretation of being the proportion of missed discoveries relative to the number of correct alternative hypotheses. Then, the missed discovery rate (MDR) of the MDF in (19) is

$$\text{MDR}(\delta^*(q; \Delta, \mathbf{A})) = E_{\mathbf{P}}\{M(\alpha^*(q; \Delta, \mathbf{A}))\}.$$

For the given δ , with proper choice of $\mathbf{A}$, we may be able to find an MDF that strongly controls the FWER or the FDR, and which possesses an optimality property with respect to another criterion, such as having a small MDR. This idea was implemented in a more restricted setting in [4] when each of the pairs of hypotheses contained simple null and simple alternative hypotheses. We point out that previous works have usually focussed in developing a particular MDF and then verifying that it controls the FWER or the FDR, such as, for example, in [3] (more comprehensively, see [27]), with notable exceptions being the papers [6,8,7,28,9,18]. It is our hope that by providing a class of MDFs where each member strongly controls the FWER, given by

$$\mathfrak{C}^\dagger = \{\delta^\dagger(q; \Delta, \mathbf{A}) : \Delta \in \mathfrak{D}, \mathbf{A} \in \mathfrak{S}\}; \quad (21)$$

or a class of MDFs where each member controls the FDR, given by

$$\mathfrak{C}^* = \{\delta^*(q; \Delta, \mathbf{A}) : \Delta \in \mathfrak{D}, \mathbf{A} \in \mathfrak{S}, \text{ with } (\Delta, \mathbf{A}) \text{ satisfying (C)}\}, \quad (22)$$

then we acquire the possibility of selecting from these classes MDFs that possess other desirable properties with respect to some suitable Type II error rate, such as the MDR. In Section 7 we provide further discussions regarding this optimality issue and present efficiency comparisons to demonstrate the viability of our idea.

5 Proofs of Theorems

The proofs of the two theorems are analogous to those of Theorem 6.1 and Theorem 7.1 in [4] which can be found in the supplemental article [29]. Note that those proofs were for special forms of the multiple decision process and multiple decision size function, whereas in the current paper we are dealing with an arbitrary element $\Delta \in \mathfrak{D}$ and an arbitrary element $\mathbf{A} \in \mathfrak{S}$. In the proofs below, we assume that $\Delta \in \mathfrak{D}$ and $\mathbf{A} \in \mathfrak{S}$ have been chosen and are fixed. Also, $q \in [0, 1]$ and recall that $\mathbf{P} \in \mathcal{P}$ denotes the true but un-known underlying probability measure. The dependence on $(\Delta, \mathbf{A}, \mathbf{P})$ of some of the relevant processes and quantities below will not be explicitly written for brevity, unless needed for clarity.

5.1 Establishing Theorem 1

Proof We start by defining the stochastic process $\mathbf{H}_1 = \{H_1(\alpha) : \alpha \in [0, 1]\}$ via

$$H_1(\alpha) \equiv H_1(\alpha; \Delta, \mathbf{A}) = \prod_{m \in \mathcal{M}} [1 - A_m(\alpha)]^{1 - \delta_m(A_m(\alpha) -)}. \quad (23)$$

The sample paths of this process are, a.e. $[\mathbf{P}]$, left-continuous with right-hand limits (*caglad*), are piecewise nonincreasing, and with

$$1 - \alpha \leq H_1(\alpha-) = H_1(\alpha) \leq H_1(\alpha+)$$

for every $\alpha \in (0, 1)$, where the first inequality is due to property (A3). In fact, by virtue of property (A1) and property (D1), note that

$$\lim_{\alpha \downarrow 0} H_1(\alpha) = 1 \text{ and } \lim_{\alpha \uparrow 1} H_1(\alpha) = 1.$$

Now, in terms of $\mathbf{H}_1$, we have that

$$\alpha^\dagger(q) = \inf\{\alpha \in [0, 1] : H_1(\alpha) < 1 - q\}.$$

Since, as pointed out above, we have $1 - \alpha \leq H_1(\alpha)$, then by its definition, we must have $\alpha^\dagger(q) \leq q$. This implies that

$$H_1(\alpha^\dagger(q)) \geq 1 - q. \quad (24)$$

For the quantity of main interest in the theorem, we have

$$E_{\mathbf{P}}[I\{S_0(\alpha^\dagger(q)) \geq 1\}] = \mathbf{P}\{S_0(\alpha^\dagger(q)) \geq 1\} = 1 - \mathbf{P}\{S_0(\alpha^\dagger(q)) = 0\} = 1 - \mathbf{P}\left\{\bigcap_{m \in \mathcal{M}_0(\mathbf{P})} [\delta_m(A_m(\alpha^\dagger(q))) = 0]\right\}.$$

The last probability cannot, however, be written as a product of probabilities since the $\delta_m(A_m(\alpha^\dagger(q)))$ for $m \in \mathcal{M}_0(\mathbf{P})$ need not be independent owing to the dependence on $\alpha^\dagger(q)$ which is determined by all the $(\alpha_m, m \in \mathcal{M})$. On the other hand, we do have the set equality

$$\bigcap_{m \in \mathcal{M}_0(\mathbf{P})} [\delta_m(A_m(\alpha^\dagger(q))) = 0] = \left\{ \alpha^\dagger(q) < \min_{m \in \mathcal{M}_0(\mathbf{P})} \alpha_m \right\}, \quad (25)$$

where the α_m s are the generalized p -value statistics defined in (12).

Next, define the stochastic process $\mathbf{H}_2 = \{H_2(\alpha) : \alpha \in [0, 1]\}$ via

$$H_2(\alpha) \equiv H_2(\alpha; \Delta, \mathbf{A}) = \left(\prod_{m \in \mathcal{M}_0(\mathbf{P})} [1 - A_m(\alpha)] \right) \left(\prod_{m \in \mathcal{M}_1(\mathbf{P})} [1 - A_m(\alpha)]^{1 - \delta_m(A_m(\alpha))} \right).$$

Analogously to the $\mathbf{H}_1$ process, this has caglad sample paths. Let us then define the quantity

$$\alpha^\#(q) \equiv \alpha^\#(q; \Delta, \mathbf{A}) = \inf\{\alpha \in [0, 1] : H_2(\alpha) < 1 - q\}.$$

Note that this is not a random variable since this depends on the unknown probability measure $\mathbf{P}$, in contrast to $\alpha^\dagger(q)$. Furthermore, also note that

$$H_2(\alpha^\#(q)) \geq 1 - q. \quad (26)$$

From their definitions, $H_1(\alpha) \leq H_2(\alpha)$, so that $H_1(\alpha) < 1 - q$ implies $H_2(\alpha) < 1 - q$. Consequently,

$$\alpha^\dagger(q) \geq \alpha^\#(q). \quad (27)$$

Now, the importance of the quantity $\alpha^\#(q)$ arises because of the crucial set equality

$$\left\{ \alpha^\dagger(q) < \min_{m \in \mathcal{M}_0(\mathbf{P})} \alpha_m \right\} = \left\{ \alpha^\#(q) < \min_{m \in \mathcal{M}_0(\mathbf{P})} \alpha_m \right\}. \quad (28)$$

To see this equality, first observe that the inclusion $\subseteq$ immediately follows from (27). To prove the reverse inclusion, since $\{\alpha^\#(q) < \min_{m \in \mathcal{M}_0(\mathbf{P})} \alpha_m\}$ implies that, for some $\alpha_0 < \min_{m \in \mathcal{M}_0(\mathbf{P})} \alpha_m$, we have $H_2(\alpha_0) < 1 - q$. But for such an α_0 , we will have $\delta_m(A_m(\alpha_0)) = 0$ for all $m \in \mathcal{M}_0(\mathbf{P})$, so that

$$\alpha_0 \in \{\alpha \in [0, 1] : H_1(\alpha) < 1 - q\}.$$

Consequently,

$$\alpha^\dagger(q) = \inf\{\alpha \in [0, 1] : H_1(\alpha) < 1 - q\} \leq \alpha_0 < \min_{m \in \mathcal{M}_0(\mathbf{P})} \alpha_m.$$

The reverse inclusion $\supseteq$ thus follows, completing the proof of (28).

By (25), (28), and the iterated expectation rule, it now follows that

$$\begin{aligned} & \mathbf{P} \left\{ \bigcap_{m \in \mathcal{M}_0(\mathbf{P})} [\delta_m(A_m(\alpha^\dagger(q))) \right. \\ & \left. = 0] \right\} = \mathbf{P} \left\{ \alpha^\dagger(q) < \min_{m \in \mathcal{M}_0(\mathbf{P})} \alpha_m \right\} = \mathbf{P} \left\{ \alpha^\#(q) < \min_{m \in \mathcal{M}_0(\mathbf{P})} \alpha_m \right\} = E_{\mathbf{P}} \left[\mathbf{P} \left\{ \alpha^\#(q) < \min_{m \in \mathcal{M}_0(\mathbf{P})} \alpha_m \mid \alpha^\#(q) \right\} \right]. \end{aligned}$$

Since $\alpha^\#(q)$ is measurable with respect to the sub- σ -field $\sigma(\delta_m : m \in \mathcal{M}_1(\mathbf{P}))$, whereas $\min_{m \in \mathcal{M}_0(\mathbf{P})} \alpha_m$ is measurable with respect to the sub- σ -field $\sigma(\delta_m : m \in \mathcal{M}_0(\mathbf{P}))$, then by condition (D3), $\alpha^\#(q)$ and $\min_{m \in \mathcal{M}_0(\mathbf{P})} \alpha_m$ are independent. Furthermore, by condition (D3), we obtain

$$\mathbf{P} \left\{ \min_{m \in \mathcal{M}_0(\mathbf{P})} \alpha_m > w \right\} = \mathbf{P} \left\{ \bigcap_{m \in \mathcal{M}_0(\mathbf{P})} [\delta_m(A_m(w)) = 0] \right\} = \prod_{m \in \mathcal{M}_0(\mathbf{P})} \mathbf{P} \{ \delta_m(A_m(w)) = 0 \} = \prod_{m \in \mathcal{M}_0(\mathbf{P})} [1 - A_m(w)],$$

with the last equality a consequence of condition (D4). Therefore,

$$\begin{aligned} & \mathbf{P} \left\{ \bigcap_{m \in \mathcal{M}_0(\mathbf{P})} [\delta_m(A_m(\alpha^\dagger(q))) \right. \\ & \left. = 0] \right\} \\ & = E_{\mathbf{P}} \left\{ \prod_{m \in \mathcal{M}_0(\mathbf{P})} [1 - A_m(\alpha^\#(q))] \right\} \geq E_{\mathbf{P}} \left\{ \left(\prod_{m \in \mathcal{M}_0(\mathbf{P})} [1 - A_m(\alpha^\#(q))] \right) \right. \\ & \quad \times \left. \left(\prod_{m \in \mathcal{M}_1(\mathbf{P})} [1 - A_m(\alpha^\#(q))]^{1 - \delta_m(A_m(\alpha^\#(q)))} \right) \right\} \\ & = E_{\mathbf{P}} \{ H_2(\alpha^\#(q)) \} \geq E_{\mathbf{P}} (1 - q) = 1 - q, \end{aligned}$$

with the last inequality following from (26). Thus, finally, we have

$$E_{\mathbf{P}} [I\{S_0(\alpha^\dagger(q)) \geq 1\}] = 1 - \mathbf{P} \left\{ \bigcap_{m \in \mathcal{M}_0(\mathbf{P})} [\delta_m(A_m(\alpha^\dagger(q))) = 0] \right\} \leq 1 - (1 - q) = q.$$

This completes the proof of Theorem 1.

We remark that condition (D4) can be weakened to just having

$$\forall m \in \mathcal{M}_0(\mathbf{P}), \forall \alpha \in [0, 1]: E_{\mathbf{P}}\{\delta_m(\alpha)\} \leq \alpha \quad (29)$$

to still get the desired strong FWER control. This is so since in the portion of the proof where we have

$$\prod_{m \in \mathcal{M}_0(\mathbf{P})} \mathbf{P}\{\delta_m(A_m(w))=0\} = \prod_{m \in \mathcal{M}_0(\mathbf{P})} [1 - A_m(w)],$$

we simply replace the second $=$ sign by $\leq$ and then the proof of the theorem goes through.

5.2 Establishing Theorem 2

Proof As we mentioned, the proof of Theorem 2 mimics that of Theorem 7.1 in [4] as presented in [29]. As an aside, we remark that the seed of the *idea* of providing a *class* of FDR-controlling multiple decision functions was planted to us upon realizing that the proof of the afore-mentioned Theorem 7.1 is functionally independent of the choice of the multiple decision size function.

The case with $q = 0$ is trivial since then $\alpha^*(0) = 0$, so that $F(\alpha^*(0)) = 0$. Thus we restrict to $q \in (0, 1]$. We first consider the case where $\mathbf{P}$ is such that $|\mathcal{M}_0(\mathbf{P})| < M$. By the defining property of $\alpha^*(q)$ given in (17), we have that

$$S(\alpha^*(q)) \geq \frac{1}{q} A_{\bullet}(\alpha^*(q)) \quad (30)$$

where $A_{\bullet}(\alpha) = \sum_{m \in \mathcal{M}_0} A_m(\alpha)$. Consequently, from (15),

$$F(\alpha^*(q)) \leq q \frac{S_0(\alpha^*(q))}{A_{\bullet}(\alpha^*(q))} I\{S(\alpha^*(q)) > 0\} \leq q \frac{S_0(\alpha^*(q))}{A_{\bullet}(\alpha^*(q))}. \quad (31)$$

For $\alpha \in [0, 1]$, define the sub- σ -field

$$\mathcal{F}_{\alpha} \equiv \mathcal{F}_{\alpha}(\mathbf{\Delta}, \mathbf{A}) = \sigma\{\delta_m(A_m(\beta)): \beta \in [\alpha, 1], m \in \mathcal{M}\}. \quad (32)$$

Observe that $\mathfrak{F} = (\mathcal{F}_{\alpha} : \alpha \in [0, 1])$ is a decreasing collection of sub- σ -fields of $\mathcal{F}$. By its definition $\alpha^*(q)$ is an $\mathfrak{F}$ -stopping time.

Let us define the process $\mathbf{T}_0 = (T_0(\alpha) : \alpha \in [0, 1])$ according to

$$T_0(\alpha) \equiv T_0(\alpha; \mathbf{\Delta}, \mathbf{A}, \mathbf{P}) = \sum_{m \in \mathcal{M}_0(\mathbf{P})} \frac{\delta_m(A_m(\alpha))}{A_m(\alpha)}.$$

Fix $0 \leq \alpha \leq \beta \leq 1$. Then, since $\delta_m \in \{0, 1\}$, we have

$$\begin{aligned}
& E_{\mathbf{P}}\{T_0(\alpha)|\mathcal{F}_\beta\} \\
&= \sum_{m \in \mathcal{M}_0(\mathbf{P})} E_{\mathbf{P}}\left\{\frac{\delta_m(A_m(\alpha))}{A_m(\alpha)}|\mathcal{F}_\beta\right\} = \sum_{m \in \mathcal{M}_0(\mathbf{P})} \left[\frac{1}{A_m(\alpha)}\right] \mathbf{P}\{\delta_m(A_m(\beta)) \\
&= 1|\mathcal{F}_\beta\} \\
&\times E_{\mathbf{P}}\{\delta_m(A_m(\alpha))|\delta_m(A_m(\beta)) \\
&= 1\} = \sum_{m \in \mathcal{M}_0(\mathbf{P})} \delta_m(A_m(\beta)) \frac{1}{A_m(\alpha)} \frac{A_m(\alpha)}{A_m(\beta)} = T_0(\beta),
\end{aligned}$$

with the last two equality only up to $\mathbf{P}$ -equivalence. The second equality follows from (D3), whereas the second-to-last equality follows since

$$E_{\mathbf{P}}\{\delta_m(A_m(\alpha))|\delta_m(A_m(\beta))=1\} = \frac{\mathbf{P}\{\delta_m(A_m(\alpha))=1, \delta_m(A_m(\beta))=1\}}{\mathbf{P}\{\delta_m(A_m(\beta))=1\}} = \frac{\mathbf{P}\{\delta_m(A_m(\alpha))=1\}}{\mathbf{P}\{\delta_m(A_m(\beta))=1\}} = \frac{A_m(\alpha)}{A_m(\beta)}$$

because of condition (A2) for the $A_m(\cdot)$ s and conditions (D2) and (D4) for the $\delta_m(\cdot)$ s. The above results show that, under $\mathbf{P}$, $\{(T_0(\alpha), \mathcal{F}_\alpha) : \alpha \in [0, 1]\}$ forms a reverse martingale process. Further, observe that $T_0(1) = |\mathcal{M}_0(\mathbf{P})|$ a.e. $[\mathbf{P}]$ due to conditions (D1) and (A1). Thus, $E_{\mathbf{P}}(T_0(1)) = |\mathcal{M}_0(\mathbf{P})|$.

From the inequality in (31), and also noting that $\alpha^*(q) \in \alpha_{(1)}$ to get the second inequality, we obtain

$$\begin{aligned}
E_{\mathbf{P}}[F(\alpha^*(q))] &\leq q E_{\mathbf{P}}\left[\frac{S_0(\alpha^*(q))}{A_{\bullet}(\alpha^*(q))}\right] = q \sum_{m \in \mathcal{M}_0(\mathbf{P})} E_{\mathbf{P}}\left[\frac{\delta_m(\alpha^*(q))}{A_{\bullet}(\alpha^*(q))}\right] = q \sum_{m \in \mathcal{M}_0(\mathbf{P})} E_{\mathbf{P}}\left[\frac{\delta_m(\alpha^*(q))}{A_m(\alpha^*(q))} \frac{A_m(\alpha^*(q))}{A_{\bullet}(\alpha^*(q))}\right] \\
&\leq q \sum_{m \in \mathcal{M}_0(\mathbf{P})} E_{\mathbf{P}}\left[\frac{\delta_m(\alpha^*(q))}{A_m(\alpha^*(q))} \left\{ \sup_{\alpha \in [\alpha_{(1)}, 1]} \frac{\max_{m \in |\mathcal{M}_0(\mathbf{P})|} A_m(\alpha)}{A_{\bullet}(\alpha)} \right\}\right] \leq q \frac{1}{|\mathcal{M}_0(\mathbf{P})|} E_{\mathbf{P}}[T_0(\alpha^*(q))] = q \frac{1}{|\mathcal{M}_0(\mathbf{P})|} E_{\mathbf{P}}[T_0(1)] \\
&= q \frac{|\mathcal{M}_0(\mathbf{P})|}{|\mathcal{M}_0(\mathbf{P})|} = q,
\end{aligned}$$

where the last inequality is obtained using condition (C), while the third-to-last equality obtains by invoking the Optional Sampling Theorem for (reverse) martingales (cf., [30]), and the second-to-last equality because of $E_{\mathbf{P}}[T_0(1)] = |\mathcal{M}_0(\mathbf{P})|$.

At this point, note that since the Šidák multiple decision size function $\mathbf{A}^S$ always satisfies condition (C) for all $\mathbf{P} \in \mathcal{P}$, including those with $\mathcal{M}_0(\mathbf{P}) = \mathcal{M}$, then $\forall \mathbf{P} \in \mathcal{D} \forall \mathbf{P} \in \mathcal{P}$, we have the property

$$E_{\mathbf{P}}\{F(\alpha^*(q; \mathbf{\Delta}, \mathbf{A}^S); \mathbf{\Delta}, \mathbf{A}^S)\} \leq q. \quad (33)$$

We now consider the case when $\mathbf{P}$ is such that $\mathcal{M}_0(\mathbf{P}) = \mathcal{M}$. Denote by $\mathcal{P}_0 = \{\mathbf{P} \in \mathcal{P} : \mathcal{M}_0(\mathbf{P}) = \mathcal{M}\}$. Consider an arbitrary $\mathbf{A} \in \mathfrak{S}$ and $\mathbf{P} \in \mathcal{P}_0$. We need to establish that $E_{\mathbf{P}}\{F(\alpha^*(q; \mathbf{A}); \mathbf{A})\} = q$. For such a $\mathbf{P} \in \mathcal{P}_0$, we have $F(\alpha; \mathbf{A}) = \mathbb{I}\{S(\alpha; \mathbf{A}) > 0\}$, so that

$$E_{\mathbf{P}}[F(\alpha^*(q; \mathbf{A}); \mathbf{A})] = \mathbf{P}\{S(\alpha^*(q; \mathbf{A}); \mathbf{A}) > 0\} = \mathbf{P}\{\alpha^*(q; \mathbf{A}) > 0\}.$$

We have, for any $\alpha \in \mathfrak{D}$ and any $\mathbf{A} \in \mathfrak{S}$ that

$$\{\alpha^*(q; \mathbf{A}) > 0\} = \bigcup_{\alpha \in (0, 1]} \left\{ \frac{S(\alpha; \mathbf{A})}{A_{\bullet}(\alpha)} \geq \frac{1}{q} \right\}. \quad (34)$$

In Lemma D.1 of [29] it was established, using an inequality of [31], that for $W_m(\eta_m), m \in \mathcal{M}$, independent Bernoulli(η_m) random variables with $\eta_m \in [0, 1]$ and satisfying $\prod_{m \in \mathcal{M}} (1 - \eta_m) = 1 - \alpha$, for each $\alpha \in [0, 1]$,

$$\mathbf{P} \left\{ \frac{\sum_{m \in \mathcal{M}} W_m(\eta_m)}{\sum_{m \in \mathcal{M}} \eta_m} \geq t \right\} \leq \mathbf{P} \left\{ \frac{\sum_{m \in \mathcal{M}} W_m(\eta_m^S)}{\sum_{m \in \mathcal{M}} \eta_m^S} \geq t \right\}, \quad (35)$$

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

Noting that, under $\mathbf{P} \in \mathcal{P}_0$, $\delta_m(A_m(\alpha))$ s are independent Bernoulli($A_m(\alpha)$), then by using the inequality in (35) and condition (A3), it follows that for $q \in (0, 1]$,

$$\mathbf{P} \left\{ \frac{S(\alpha; \mathbf{A})}{A_{\bullet}(\alpha)} \geq \frac{1}{q} \right\} \leq \mathbf{P} \left\{ \frac{S(\alpha; \mathbf{A}^{S+})}{A_{\bullet}^{S+}} \geq \frac{1}{q} \right\}, \quad (36)$$

where the Šidák sizes $\mathbf{A}^{S+} = (A_m^{S+}, m \in \mathcal{M})$ in (36) have components

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

with α^+ satisfying $\prod_{m \in \mathcal{M}} [1 - A_m(\alpha)] = 1 - \alpha^+$. Observe that by (A3), we have necessarily that $\alpha^+ \leq \alpha$. Combining the results in (34) and (36), we obtain

$$\mathbf{P}\{\alpha^*(q; \mathbf{A}) > 0\} \leq \mathbf{P}\{\alpha^*(q; \mathbf{A}^{S+}) > 0\}.$$

But since we have already established that, for $\mathbf{P} \in \mathcal{P}_0$, we have

$$\mathbf{P}\{\alpha^*(q; \mathbf{A}^{S+}) > 0\} \leq q,$$

then it follows that $\mathbf{P}\{\alpha^*(q; \mathbf{A}) > 0\} \leq q$. This implies finally that

$$E_{\mathbf{P}}\{F(\alpha^*(q; \mathbf{A}); \mathbf{A})\} \leq q$$

for any $P \in \mathcal{P}_0$. This completes the proof of Theorem 2.

Short and neat proofs of FDR control for multiple decision functions based on p -value statistics are also provided in Theorem 4.1 of [18] and in Proposition 2.7 of [8]. The two sufficient conditions in the latter paper do not cover our set-up since the size functions need not belong to their factorized threshold collection. In addition, their dependency control condition appear to be quite distinct from our condition (C). It is not clear to us whether similar short proofs could be employed in establishing our Theorem 2 using the representation of $\delta^*(q)$ in terms of the generalized p -value statistics given in Section 6. Also, in contrast to Theorem 1 where we were able to have the weaker version of condition (D4) given in (29), we could not do this for Theorem 2. The reason is that we could *not* conclude under this weaker condition that the process $\{(T_0(\alpha), \mathcal{T}_\alpha) : \alpha \in [0, 1]\}$ is a reverse supermartingale, which would have allowed us to get the desired result. It may be possible that under certain situations we do have this supermartingale property, but the weaker condition (29) appears not sufficient for this property to hold in general.

6 Representations of MDFs in Terms of the Generalized p -Values

This section expresses the MDFs $\delta^\dagger(q)$ in (18) and $\delta^*(q)$ in (19) in terms of the generalized p -value statistics defined in (12). Without loss of generality, let $\mathcal{M} = \{1, 2, \dots, M\}$. Define the vector of anti-rank statistics via

$$((1), (2), \dots, (M)) : (\mathcal{X}, \mathcal{F}) \rightarrow (\mathfrak{M}, \sigma(\mathfrak{M})) \quad (37)$$

where $\mathfrak{M}$ is the space of all possible permutations of $\mathcal{M}$, and such that

$$\alpha_{(1)} < \alpha_{(2)} < \dots < \alpha_{(M)}.$$

Let us first consider the random variable $\alpha^\dagger(q)$ in (16). We see from its definition and those of the generalized p -value statistics that, for some $J \in \mathcal{M} \equiv \{0\} \cup \mathcal{M}$, we have $\alpha^\dagger(q) \in [\alpha_{(J)}, \delta_{(J+1)})$ if and only if

$$\begin{aligned} \forall j \in \{1, 2, \dots, J\} : \prod_{m \in \mathcal{M}} [1 - A_{(m)}(\alpha_{(j)})]^{1 - \delta_{(m)}[A_{(m)}(\alpha_{(j)}) -]} &\geq 1 - q; \\ \prod_{m \in \mathcal{M}} [1 - A_{(m)}(\alpha_{(J+1)})]^{1 - \delta_{(m)}[A_{(m)}(\alpha_{(J+1)}) -]} &< 1 - q. \end{aligned}$$

From the definition of the generalized p -value statistics we further have

$$\delta_{(m)}[A_{(m)}(\alpha_{(j)}) -] = I\{m \leq j - 1\}.$$

Consequently, by defining the $\mathcal{M}$ -valued random variable

$$J^\dagger(q) = \max \left\{ k \in \mathcal{M} : \prod_{m=j}^M [1 - A_{(m)}(\alpha_{(j)})] \geq 1 - q, j=1, 2, \dots, k \right\}, \quad (38)$$

we have the result that $\alpha^\dagger(q) \in [\alpha_{(J^\dagger(q))}, \alpha_{(J^\dagger(q)+1)}]$. As a consequence we obtain the representation of $\delta^\dagger(q)$ in (18) in terms of the α_m s given by

$$\delta^\dagger(q) \equiv (\delta_m(A_m(\alpha^\dagger(q))), m \in \mathcal{M}) = (\delta_m(A_m(\alpha_{(J^\dagger(q))})), m \in \mathcal{M}), \quad (39)$$

where we used the fact that, for each $m \in \mathcal{M}$, δ_m is constant in the interval

$$[A_m(\alpha_{(J^\dagger(q))}), A_m(\alpha_{(J^\dagger(q)+1)})].$$

Next let us consider the random variable $\alpha^*(q)$ in (17). We may re-express its defining equation via

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

But, since $\sum_{m \in \mathcal{M}} \delta_{(m)}[A_{(m)}(\alpha_{(j)})] = j$, then $\alpha^*(q) \in [\alpha_{(j)}, \alpha_{(j+1)}]$ iff

$$\sum_{m \in \mathcal{M}} A_{(m)}(\alpha_{(j)}) \leq qj; \forall j \in \{J+1, J+2, \dots, M\}; \sum_{m \in \mathcal{M}} A_{(m)}(\alpha_{(j)}) > qj.$$

Defining the $\mathcal{M}$ -valued random variable

$$J^*(q) = \max \left\{ k \in \mathcal{M}: \sum_{m \in \mathcal{M}} A_{(m)}(\alpha_{(k)}) \leq qk \right\}, \quad (40)$$

we then have that $\alpha^*(q) \in [\alpha_{(J^*(q))}, \alpha_{(J^*(q)+1)}]$. As a consequence, an equivalent representation of the MDF $\delta^*(q)$ in (19) in terms of the α_m s is provided by

$$\delta^*(q) \equiv (\delta_m(A_m(\alpha^*(q))), m \in \mathcal{M}) = (\delta_m(A_m(\alpha_{(J^*(q))})), m \in \mathcal{M}). \quad (41)$$

The representations in (39) for $\delta^\dagger(q)$ and (41) for $\delta^*(q)$ provide alternative computational approaches since, instead of computing $\alpha^\dagger(q)$ and $\alpha^*(q)$, we may simply compute the generalized p -values, then $J^\dagger(q)$ and $J^*(q)$, and then finally the realizations of the decision functions.

For a simple application, let us see what becomes of the MDFs $\delta^\dagger(q)$ and $\delta^*(q)$ if we use the Šidák multiple decision size function A^S given in (6). We use the alternate representations just obtained above. By simple manipulations, we immediately obtain that

$$J^\dagger(q) = \max \{ k \in \mathcal{M}: \alpha_{(j)} \leq 1 - (1 - q)^{M/(M-j+1)}, j=1, 2, \dots, k \}; \quad J^*(q) = \max \{ k \in \mathcal{M}: M[1 - (1 - \alpha_{(k)})]^{1/M} \leq qk \}.$$

But, for these Šidák size functions, the (ordinary) p -value statistics are given by

$$P_m = A_m^S(\alpha_m) = 1 - (1 - \alpha_m)^{1/M}, m \in \mathcal{M}.$$

Re-expressing the $J^\dagger(q)$ and $J^*(q)$ in terms of these p -values, we easily obtain by simple manipulations that

$$J^\dagger(q) = \max\{k \in \mathcal{M} : P_{(j)} \leq 1 - (1 - q)^{1/(M-j+1)}, j=1, 2, \dots, k\}; \quad (42)$$

$$J^*(q) = \max\{k \in \mathcal{M} : P_{(k)} \leq qk/M\}. \quad (43)$$

Observe that $J^\dagger(q)$ in (42) leads to the step-down sequential Šidák FWER-controlling procedure, see [10,27]; whereas, $J^*(q)$ in (43) is the usual form of the step-up Benjamini-Hochberg FDR-controlling procedure in [3]. Thus, through the Šidák sizes, we are able to obtain from our formulation two popular MDFs for FWER and FDR control as special cases of the MDFs $\delta^\dagger(q)$ and $\delta^*(q)$!

7 On Developing Optimal MDFs

7.1 Computing Optimal Size Functions

In this subsection, we demonstrate the potential utility of the classes of MDFs arising from Theorems 1 and 2 in the context of obtaining MDFs with some optimality properties, especially in non-exchangeable multiple hypotheses testing settings, which include those where the power characteristics of the M test functions are not identical.

Consider a fixed multiple decision process $\mathfrak{D} \in \mathfrak{D}$ and fix a probability measure $\mathbf{P}_1 \in \mathcal{P}$, with $\mathcal{M}_1(\mathbf{P}_1) = \mathcal{M}$. Recall that in the single hypothesis testing situation, focus on the power of a test procedure is usually on a specific alternative determined by a specified effect size ideally set at the design of experiment stage so as to determine an appropriate sample size. The fixed $\mathbf{P}_1$ is tantamount to specifying the alternative $\mathbf{P}$ that is of most interest.

Define the mappings $\pi_m : [0, 1] \rightarrow [0, 1]$ for $m \in \mathcal{M}$ according to

$$\pi_m(\alpha; \mathbf{P}_1) = E_{\mathbf{P}_1}[\delta_m(\alpha)], \alpha \in [0, 1]. \quad (44)$$

When viewed as a function of $\mathbf{P}_1$, $\pi_m(\alpha; \cdot)$ is the power function of δ_m when it is allocated a size of α . Of interest to us, though, is to view it as a function of α for the fixed $\mathbf{P}_1$. In this case, $\pi_m(\cdot; \mathbf{P}_1)$ is the receiver operating characteristic curve (ROC) of the m th test or decision process. Assume that for each $m \in \mathcal{M}$, the mapping $\alpha \mapsto \pi_m(\alpha; \mathbf{P}_1)$ is strictly increasing with $\pi_m(1; \mathbf{P}_1) = 1$ and twice-differentiable.

Suppose we desire to strongly control the overall FWER or FDR at some pre-specified level $q \in [0, 1]$, but at the same time maximize the total (or average) power at $\mathbf{P} = \mathbf{P}_1$. Our idea is to first obtain the optimal multiple decision size function for *weak* FWER control associated with $\mathfrak{D}$, denoted by $\mathbf{A}^* = (A_m^*(\alpha), m \in \mathcal{M}) \in \mathfrak{A}$. This is the multiple decision size function $\mathbf{A}$ satisfying the condition $\forall \alpha \in [0, 1] : \prod_{m \in \mathcal{M}} [1 - A_m(\alpha)] = 1 - \alpha$, and such that the total

power at $\mathbf{P} = \mathbf{P}_1$, given by $\max_{m \in \mathcal{M}} \pi_m(A_m(\alpha); \mathbf{P}_1)$, is maximized. We formally present this optimal multiple decision size function in the following theorem.

Theorem 3 Assume that for the multiple decision process $\pi_m = (\pi_m : m \in \mathcal{M})$, each of the M ROC functions $\alpha \mapsto \pi_m(\alpha; \mathbf{P}_1)$ are strictly increasing, concave, and twice-differentiable (in α) with first and second derivatives $\pi'_m(\cdot; \mathbf{P}_1)$ and $\pi''_m(\cdot; \mathbf{P}_1)$, respectively, so that $\pi''_m(\alpha; \mathbf{P}_1) < 0$. Then the multiple decision size function $\mathbf{A} = (A_m(\alpha) : m \in \mathcal{M})$ maximizing the global power of the multiple decision function $\delta(\alpha) = (\delta_m[A_m(\alpha)] : m \in \mathcal{M})$ under weak FWER-control at level α is $\mathbf{A}^*(\Delta) = (A_m^*(\alpha) : m \in \mathcal{M})$ satisfying the two sets of conditions

i. For some $\lambda \in \mathbb{R}_+$ and for each $m \in \mathcal{M}$,

$$\log\{\pi'_m[A_m^*(\alpha); \mathbf{P}_1]\} + \log[1 - A_m^*(\alpha)] = \log(\lambda);$$

ii. $\sum_{m \in \mathcal{M}} \log[1 - A_m^*(\alpha)] = \log(1 - \alpha)$.

Equivalently, for fixed $\alpha \in (0, 1)$, $A_m^*(\alpha) = z_m$, $m \in \mathcal{M}$, with $\mathbf{z} = (z_1, \dots, z_M)^t$ satisfying the set of M equations

$$\left[\mathbf{I} - \frac{1}{M} \mathbf{J} \right] \kappa(\mathbf{z}) + \mathbf{l}(\mathbf{z}) - \left[\frac{1}{M} \log(1 - \alpha) \right] \mathbf{1} = \mathbf{0}$$

where $\kappa(\mathbf{z}) = (\log \pi'_m(z_m; \mathbf{P}_1), m=1, 2, \dots, M)^t$, $\mathbf{l}(\mathbf{z}) = (\log(1 - z_m), m=1, 2, \dots, M)^t$, $\mathbf{I}$ is the $M \times M$ identity matrix, $\mathbf{J} = \mathbf{1}\mathbf{1}^t$, and $\mathbf{1}$ ($\mathbf{0}$) is the $M \times 1$ vector of 1's (0's).

Proof The proof of this theorem is straightforward by using Lagrangian optimization which leads to the two conditions (i) and (ii). To obtain the equivalent set of equations, sum the M elements in item (i), use the constraint in item (ii), plug-in expression for $\log(\lambda)$ into the equations in item (i), and finally convert into vectors and matrices. The concavity of the ROC functions guarantees that a maximizer is obtained.

Conditions (i) and (ii) in Theorem 3 are analogous to the conditions in Theorem 4.3 in [4] which dealt with the situation when the individual test functions coincide with the Neyman-Pearson most powerful tests. The concavity of the ROC functions is critical and guarantees that these functions are above the 45-degree line, which implies that each $\delta_m(\alpha)$ is an unbiased test function, that is, performs better than the test function which does not use the data x_m but always rejects H_{m0} with probability α . The alternative set of equations in Theorem 3 is conducive to a Newton-Raphson iteration for finding the (approximate) solution $\mathbf{z}$. To perform this iterative approach, it is computationally more stable to perform the iterations on the $\mathbf{v} = (v_1, v_2, \dots, v_M)^t$ with

$$v_m \equiv v_m(z_m) = \log \left[\frac{z_m}{1 - z_m} \right], m=1, 2, \dots, M.$$

Thus, $z_m \equiv z_m(v_m) = \exp(v_m)/(1 + \exp(v_m))$ for $m = 1, 2, \dots, M$. Define the mappings

$$\mathbf{V}(\mathbf{v}) = [\mathbf{I} - \frac{1}{M} \mathbf{J}] \kappa(\mathbf{z}(\mathbf{v})) + \mathbf{l}(\mathbf{z}(\mathbf{v})) - [\frac{1}{M} \log(1 - \alpha)] \mathbf{1}; \mathbf{H}(\mathbf{v}) = [\mathbf{I} - \frac{1}{M} \mathbf{J}] \mathbf{D}_1(\mathbf{v}) + \mathbf{D}_2(\mathbf{v}),$$

where the matrices $\mathbf{D}_1(\mathbf{v})$ and $\mathbf{D}_2(\mathbf{v})$ are defined, respectively, via

$$\mathbf{D}_1(\mathbf{v}) = \text{Dg} \left(\frac{\pi_m''(z_m)}{\pi_m'(z_m)} z_m (1 - z_m), m=1, 2, \dots, M \right); \mathbf{D}_2(\mathbf{v}) = -\text{Dg}(z_m, m=1, 2, \dots, M).$$

In the above expressions, $\text{Dg}(w_1, w_2, \dots, w_M)$ means the diagonal matrix with elements $w_1, w_2, \dots, w_M$. The Newton-Raphson iteration updating for $\mathbf{v}$ then proceeds via

$$\mathbf{v}^{\text{new}} \leftarrow \mathbf{v}^{\text{old}} - [\mathbf{H}(\mathbf{v}^{\text{old}})]^{-1} [\mathbf{V}(\mathbf{v}^{\text{old}})].$$

Seed values for the v_m 's are those associated with the Šidák sizes $z_m = 1 - (1 - \alpha)^{1/M}$ for $m = 1, 2, \dots, M$. A possible computational bottleneck in this approach is the inversion of the potentially huge matrix $\mathbf{H}$. If this becomes problematic, then one may revert to directly using conditions (i) and (ii) in Theorem 3 as was done in the computational portion in [4].

Having determined the optimal multiple decision size function $\mathbf{A}^*$ associated with $\mathbf{P}_1$, which is at this point optimal only in the sense of *weak* FWER control, we can then apply Theorem 1 to obtain the MDF $\delta^\dagger(q; \mathbf{P}_1, \mathbf{A}^*)$ which will *strongly* control the FWER at q ; or apply Theorem 2 to obtain the MDF $\delta^*(q; \mathbf{P}_1, \mathbf{A}^*)$ which will (strongly) control the FDR at q . By virtue of the choice of the size process $\mathbf{A}^*$, which is tied-in to the multiple decision process and the target probability measure $\mathbf{P}_1$, we expect that the MDFs $\delta^\dagger(q; \mathbf{P}_1, \mathbf{A}^*)$ and $\delta^*(q; \mathbf{P}_1, \mathbf{A}^*)$ will perform better with respect to overall power at $\mathbf{P}_1$ relative to, for example, the sequential Šidák MDF or the BH MDF, which we saw from the preceding section are MDFs arising from the Šidák multiple decision size function, a size function that may not be optimal for the chosen multiple decision size process $\mathbf{P}_1$.

7.2 Results of a Modest Simulation Study

To demonstrate the improvement in performance of the proposed procedures, we provide results of a simulation study showing the gain in global power, or the decrease in the missed discovery rate, of the proposed MDF. We focus only on the FDR-controlling procedure in this simulation, and compare its performance with the BH procedure. Note that results of a simulation study was presented in [4] demonstrating the improvement over the BH procedure of the MDF δ^* in a Gaussian setting. In the current simulation study, we consider a non-Gaussian model.

The model considered in the simulation study has $T_m \sim Ga(n_m, 1/\theta_m)$ for $m = 1, 2, \dots, M$, so that $E[T_m] = n\theta_m$. We assumed that the T_m 's are independent random variables. The multiple decision problem is to decide, for each m , whether $H_{m0} : \theta_m = \theta_{m0}$ or $H_{m1} : \theta_m > \theta_{m0}$. Denote by $\mathcal{C}(\cdot; k)$, $c(\cdot; k)$, and $\mathcal{C}^{-1}(\cdot; k)$ the distribution, density, and quantile functions, respectively, of a chi-squared random variable with degrees-of-freedom k . The uniformly

most powerful (UMP) decision process for testing H_{m0} versus H_{m1} using only T_m is given by $\delta_m = (\delta_m(\cdot; \alpha) : \alpha \in [0, 1])$ where

$$\delta_m(t_m; \alpha) = I\{2t_m/\theta_{m0} \geq \mathcal{C}^{-1}(1 - \alpha; 2n_m)\}.$$

The associated P -value statistic for this test process is

$$P_m(T_m) = 1 - \mathcal{C}(2T_m/\theta_{m0}; 2n_m),$$

so the decision function $\delta_m(t_m; \alpha)$ could also be expressed via

$$\delta_m(t_m; \alpha) = I\{P_m(t_m) \leq \alpha\}.$$

For a fixed θ_{m1} , which exceeds θ_{m0} , the ROC function associated with the decision process δ_m is easily seen to be

$$\pi_m(\alpha; \theta_{m0}, \theta_{m1}) = \pi_m(\alpha; \xi_m) = 1 - \mathcal{C}(\xi_m \mathcal{C}^{-1}(1 - \alpha; 2n_m); 2n_m),$$

where $\xi_m = \theta_{m0}/\theta_{m1}$, which can be viewed as the effect size, though it might be more fitting to call this as the reciprocal of the *usual* effect size. It is then straight-forward to verify that

$$\pi'_m(\alpha; \xi_m) = \xi_m^{n_m} \exp\left\{-\frac{1}{2}(\xi_m - 1)\mathcal{C}^{-1}(1 - \alpha; 2n_m)\right\}; \quad \pi''_m(\alpha; \xi_m) = [(\xi_m - 1)/2]\pi'_m(\alpha; \xi_m)/c[\mathcal{C}^{-1}(1 - \alpha; 2n_m); 2n_m].$$

These are the functions that are needed to implement the Newton-Raphson iterative procedure described in the preceding subsection for finding the optimal size functions for weak-FWER control, and consequently for finding $\alpha^*(q)$. We implemented this algorithm using an R program and with use of the `pchisq`, `dchisq`, and `qchisq` objects. A computational limitation encountered is when $A_m(\alpha)$ becomes very, very small which causes `qchisq` to output either an `Inf` or an `NaN`. When this is encountered for a generated sample, the sample is discarded from the simulation. In the simulation runs, this limitation arose in Table 1 only for the case with $(n, p) = (30, .50)$. For this case, 120 samples got discarded prior to obtaining 1000 successful replications.

In the simulation study, we fixed $M = 100$ and $n_m = n \in \{5, 10, 30\}$ for $m = 1, 2, \dots, M$. We then specified $p \in \{.2, .3, .5\}$, which is the theoretical proportion of correct H_{m1} 's. The basic simulation experiment goes as follows: We generated $B_1, B_2, \dots, B_M$ to be IID $\text{Ber}(p)$. We then generated $\xi_1, \xi_2, \dots, \xi_M$ to be IID from a uniform distribution on $[\cdot 25, \cdot 75]$. [We note that this step is typically outside the basic simulation experiment, but we wanted to explore many effect sizes in the study.] With $\theta_{m0} = 1$, $m = 1, 2, \dots, M$, the true scale parameters were computed to be $\theta_m = (1 - B_m)\theta_{m0} + B_m\theta_{m0}/\xi_m$. We then generate the data $T_m \sim \text{Ga}(n_m, 1/\theta_m)$ for $m = 1, 2, \dots, M$. For the observed data $\mathbf{t} = (t_1, t_2, \dots, t_M)$, and for an FDR threshold of $q = .20$, we applied the BH procedure and the FDR-controlling procedure $\delta^*(q)$, and for each procedure we computed the false discovery proportion (FDP) and the missed discovery

proportion (MDP). Recall that for an action vector $\mathbf{a} = (a_1, a_2, \dots, a_M) \in \{0, 1\}^M$, the FDP and MDP are the quantities

$$FDP = \frac{\sum_{m=1}^M a_m(1 - B_m)}{\max(1, \sum_{m=1}^M a_m)} \text{ and } MDP = \frac{\sum_{m=1}^M (1 - a_m)B_m}{\max(1, \sum_{m=1}^M B_m)}.$$

This basic experiment was replicated 1000 times, and the estimated FDR and MDR (missed discovery rate) are the averages of the FDP's and MDP's over these 1000 replications. Table 1 presents the results of this simulation study. From this table we see that both the BH and $\delta^*(q)$ procedures control their FDR rates to be no more than the pre-specified q . We also observe that for the simulation runs performed, there is a slight improvement in terms of the missed discovery rate of the procedure $\delta^*(q)$ over the BH procedure, with the improvement increasing as n increases. At the same time, it should be noted that the BH procedure, even though it does *not* exploit the ROC differences of the individual UMP decision processes, still performs quite comparably well in these simulation runs, lending credence to its practical appeal and utility owing to its simplicity and ease of implementation! We mentioned above that for the case $(n, p) = (50, .50)$, some samples were discarded due to a computational limitation. We cannot conclude whether this led to some bias in the results. However, even without this case, the improvement of the proposed procedure over the BH is evident based on all the other cases in Table 1.

Finally, we performed one more set of simulation runs with $M = 100$, $n \in \{5, 10, 30\}$, $p = .2$, but with the effect sizes taking only four different values. The effect size vector is $\xi = (\xi_m, m = 1, 2, \dots, M)$ with $\xi_m = .1$, $m = 1, 2, \dots, 25$, $\xi_m = .3$, $m = 26, 27, \dots, 50$, $\xi_m = .5$, $m = 51, 52, \dots, 75$, and $\xi_m = .7$, $m = 76, 77, \dots, 100$. The number of replications is again 1000. The results of these simulation runs are presented in Table 2. From this table we observe a more pronounced advantage of $\delta^*(q)$ over the BH procedure in terms of their MDRs, especially when n is equal to 30. Both procedures still satisfied the FDR-threshold of $q = 20\%$.

Finally, note from the results of these simulation studies that these procedures do not yet exhaust the pre-specified FDR-threshold. This indicates that some improvements in global power could still be made as what has been done for the BH procedure by incorporating an adjustment on the threshold using an estimate of the proportion of correct alternative hypotheses. However, we do not explore possible similar adjustments to our proposed procedures in the current paper. We refer the reader to [18] for more discussion on this α -exhaustion notion.

8 Concluding Remarks

In this paper we provided classes of FWER-controlling and FDR-controlling procedures derived from individual decision processes. The innovative idea is the use of multiple decision size functions, which serve as size-pickers for each of the decision processes. Through these classes of MDFs, optimal size functions may be obtained with respect to global Type II error at a specified alternative hypothesis probability measure $\mathbf{P}_1$, thereby potentially identifying MDFs with some optimal properties. Via a simulation study, the

FDR-controlling procedure was demonstrated to perform better than the BH procedure for gamma-distributed data. Other issues have not been addressed which maybe of interest in future work. For instance, it would be desirable to extend the results in settings where the components of $\{\delta_m : m \in \mathcal{M}_0(\mathbf{P})\}$ are dependent as in [32,33]; see also the review paper [34]. Another potential extension is to consider generalized FWER and FDR as in [35]. A possible criticism of the proposed approach is the need to know the ROC functions, which thereby entails the specification of a fixed $\mathbf{P}_1$, or equivalently, effect sizes for each of the M testing problems. However, this is similar to what is typically done even with single hypothesis testing problems where we focus on a specific alternative hypothesis which is of most interest to detect, such as in the sample size determination problem. In future work we are planning on exploring the use of *asymptotic* ROC functions which will arise by letting the sample sizes (the n_m 's in the concrete example) and M increase and then considering a contiguous set of alternatives, e.g., Pitman-type alternatives. This may potentially alleviate or lessen the criticism mentioned above. Certainly, another approach to eliminating the need to specify a fixed $\mathbf{P}_1$ to obtain the ROC functions is to use a Bayesian approach. This programme was partly done in [24], though more work is still needed to clarify and fully understand this approach.

Acknowledgments

We thank Professor Sanat Sarkar for helpful discussions and sincerely thank the reviewers of this work for their sharp and critical comments which were extremely helpful in improving the manuscript. We very much thank *Metrika* editors, Professor Norbert Henze and Professor Udo Kamps, for providing an outlet for this work.

The authors acknowledge support from National Science Foundation (NSF) Grants DMS 0805809 and DMS 1106435, National Institutes of Health (NIH) Grants P20RR17698, R01CA154731, and P30GM103336-01A1.

References

1. Efron B. Microarrays, empirical Bayes and the two-groups model. *Statist. Sci.* 2008; 23(1):1–22.
2. Efron B. Simultaneous inference: When should hypothesis testing problems be combined? *The Annals of Applied Statistics.* 2008; 1:197–223.
3. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B.* 1995; 57(1):289–300.
4. Peña EA, Habiger JD, Wu W. Power-enhanced multiple decision functions controlling family-wise error and false discovery rates. *Ann. Statist.* 2011; 39(1):556–583.
5. Westfall PH, Krishen A. Optimally weighted, fixed sequence and gatekeeper multiple testing procedures. *J. Statist. Plann. Inference.* 2001; 99(1):25–40.
6. Genovese CR, Roeder K, Wasserman L. False discovery control with p -value weighting. *Biometrika.* 2006; 93(3):509–524.
7. Roeder K, Wasserman L. Genome-wide significance levels and weighted hypothesis testing. *Statist. Sci.* 2009; 24(4):398–413.
8. Blanchard G, Roquain E. Two simple sufficient conditions for FDR control. *Electron. J. Stat.* 2008; 2:963–992.
9. Roquain E, van de Wiel MA. Optimal weighting for false discovery rate control. *Electronic Journal of Statistics.* 2009; 3:678–711.
10. Holm S. A simple sequentially rejective multiple test procedure. *Scand. J. Statist.* 1979; 6(2):65–70.
11. Goeman JJ, Solari A. The sequential rejection principle of familywise error control. *Ann. Statist.* 2010; 38(6):3782–3810.

12. Westfall PH, Krishen A, Young SS. Using prior information to allocate significance levels for multiple endpoints. *Statistics in Medicine*. 1998; 17:2107–2119. [PubMed: 9789917]
13. Genovese, C.; Wasserman, L. *Bayesian statistics*, 7 (Tenerife, 2002). New York: Oxford Univ. Press; 2003. Bayesian and frequentist multiple testing; p. 145-161. With discussions by Merlise A. Clyde and Christian P. Robert and Judith Rousseau, and a reply by the authors
14. Storey J. The optimal discovery procedure: a new approach to simultaneous significance testing. *Journal of the Royal Statistical Society, Series B*. 2007; 69:347–368.
15. Sun W, Cai T. Oracle and adaptive compound decision rules for false discovery rate control. *Journal of the American Statistical Association*. 2007; 102:901–912.
16. Sarkar SK, Zhou T, Ghosh D. A general decision theoretic formulation of procedures controlling FDR and FNR from a Bayesian perspective. *Statist. Sinica*. 2008; 18(3):925–945.
17. Kang G, Ye K, Liu N, Allison D, Gao G. Weighted multiple hypothesis testing procedures. *Statistical Applications in Genetics and Molecular Biology*. 2009; 8:1–21.
18. Finner H, Dickhaus T, Roters M. On the false discovery rate and an asymptotically optimal rejection curve. *Ann. Statist.* 2009; 37(2):596–618.
19. Habiger JD, Peña EA. Compound p -value statistics for multiple testing procedures. *J. Multivariate Anal.* 2014; 126:153–166.
20. Müller P, Parmigiani G, Robert C, Rousseau J. Optimal sample size for multiple testing: the case of gene expression microarrays. *J. Amer. Statist. Assoc.* 2004; 99(468):990–1001.
21. Scott J, Berger J. An exploration of aspects of Bayesian multiple testing. *Journal of Statistical Planning and Inference*. 2006; 136:2144–2162.
22. Guindani M, Muller P, Zhang S. A Bayesian discovery procedure. *JRSS B*. 2009; 71:905–925.
23. Bogdan M, Chakrabarti A, Frommlet F, Ghosh JK. Asymptotic Bayes-optimality under sparsity of some multiple testing procedures. *Ann. Statist.* 2011; 39(3):1551–1579.
24. Wu W, Peña EA. Bayes multiple decision functions. *Electron. J. Stat.* 2013; 7:1272–1300. [PubMed: 25414762]
25. Habiger JD, Peña EA. Randomized P -values and nonparametric procedures in multiple testing. *J. Nonparametr. Stat.* 2011; 23(3):583–604. [PubMed: 25419090]
26. Šidák Z. Rectangular confidence regions for the means of multivariate normal distributions. *J. Amer. Statist. Assoc.* 1967; 62:626–633.
27. Dudoit, S.; van der Laan, MJ. *Springer Series in Statistics*. New York: Springer; 2008. Multiple testing procedures with applications to genomics.
28. Blanchard G, Roquain É. Adaptive false discovery rate control under independence and dependence. *J. Mach. Learn. Res.* 2009; 10:2837–2871.
29. Peña EA, Habiger JD, Wu W. Supplement to ‘Power-Enhanced Multiple Decision Functions Controlling Family-Wise Error and False Discovery Rates’. 2011
30. Doob, JL. *Stochastic processes*. New York: John Wiley & Sons Inc.; 1953.
31. Hoeffding W. On the distribution of the number of successes in independent trials. *Ann. Math. Statist.* 1956; 27:713–721.
32. Sarkar SK, Chang CK. The Simes method for multiple hypothesis testing with positively dependent test statistics. *J. Amer. Statist. Assoc.* 1997; 92(440):1601–1608.
33. Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.* 2001; 29(4):1165–1188.
34. Sarkar SK. On methods controlling the false discovery rate. *Sankhyā*. 2008; 70(2, Ser. A):135–168.
35. Sarkar SK. Stepup procedures controlling generalized FWER and generalized FDR. *Ann. Statist.* 2007; 35(6):2405–2420.

Fig. 1.

Demonstration of the optimal size functions in the restricted family of size functions for effect sizes $\gamma = (1.08, 1.43, 0.19, 1.73, 0.10)$ with $M = 5$. The first plot frame contains the plots of the $\alpha \mapsto b_m(\alpha)$ functions, the second plot frame is for the $\alpha \mapsto A_m(\alpha) = 1 - (1 - \alpha)^{k_m(\alpha)}$ functions, and the third plot frame is for the $\alpha \mapsto k_m(\alpha)$ functions.

Table 1

Results of simulation study for gamma-distributed component data. The effects sizes were generated according to $\xi_m \sim U[.25, .75]$ for $m = 1, 2, \dots, M$. The FDR threshold was set to $q = 20\%$. The theoretical proportion of correct H_{m1} 's is p . The number of simulation replications for each combination of simulation parameters was 1000. Note that the estimated FDRs and MDRs are in percentages.

M	$n_m = n$	p	BH		$\delta^*(q)$		
			FDR	MDR	FDR	MDR	MDR
100	5	.2	15.93574	55.56008	16.03789		55.24426
	10	.2	15.78390	38.48281	14.97451		38.12848
	30	.2	16.09659	10.88535	14.93114		8.026157
100	5	.3	13.42253	46.06943	13.24371		45.90983
	10	.3	13.76047	31.97512	13.65171		30.47509
	30	.3	13.80077	5.639566	13.89979		2.626881
100	5	.5	10.37083	41.65425	10.06751		41.40670
	10	.5	10.11820	26.34404	10.11692		23.57873
	30	.5	10.03628	7.615538	9.590189		4.151696

Simulation results for the case where the effect sizes only takes 4 distinct values, and with $M = 100$, $p = .2$, $n \in \{5, 10, 30\}$. The number of replications is 1000 and the FDR threshold is $q = 20\%$.

Table 2

M	$n_m = n$	p	BH		$\delta^*(q)$		
			FDR	MDR	FDR	MDR	MDR
100	5	.2	16.3581	39.05525	15.79368	36.95671	
100	10	.2	15.86433	26.73459	16.05328	22.97558	
100	30	.2	16.0207	9.896752	15.76027	5.316628	