Classes of Multiple Decision Functions Strongly Controlling FWER and FDR

Edsel A. Pena, Joshua Habiger,
Wensong Wu

Technical Report #2009-6
July 7, 2009

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

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

Classes of Multiple Decision Functions Strongly Controlling FWER and FDR

Edsel A. Peña* Joshua Habiger† Wensong Wu‡

July 7, 2009

Abstract

This paper provides 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). These classes offer the possibility that an optimal multiple decision function with respect to a pre-specified criterion, such as the missed discovery rate (MDR), could be found within these classes. Such multiple decision functions can be utilized in multiple testing, specifically, but not limited to, the analysis of "large M, small n" microarray data sets.

Keywords and Phrases: False discovery rate; family-wise error rate; missed discovery rate; multiple decision problem; multiple testing; strong control.

1 Mathematical Setting

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 $(\mathcal{X}, \mathcal{F})$. In decision problems with action space $\mathfrak{A} = \{0, 1\}$, such as in hypothesis testing, a nonrandomized decision function is a $\delta : (\mathcal{X}, \mathcal{F}) \to (\mathfrak{A}, \sigma(\mathfrak{A}))$. In the hypothesis testing setting, a decision $\delta = 0$ corresponds to deciding in favor of a null hypothesis (H_0), whereas a decision of $\delta = 1$, usually called a *discovery*, corresponds to rejecting H_0 in favor of an alternative hypothesis (H_1). The restriction to nonrandomized decision functions does not limit the generality of the results since if we have a randomized

*E. Peña (pena@stat.sc.edu) is Professor, Department of Statistics, University of South Carolina, Columbia, SC 29208.

†J. Habiger (habigerj@stat.sc.edu) is a PhD student, Department of Statistics, University of South Carolina, Columbia, SC 29208.

‡W. Wu (wu26@math.sc.edu) is a PhD student, Department of Statistics, University of South Carolina, Columbia, SC 29208.

decision function, then through the use of a randomizer, e.g., a uniform random variate, we could convert such a decision function to a nonrandomized one. Thus, the sample space $\mathcal{X}$ may in practice be the product space between the data space and $[0,1]$. This may for instance be needed when dealing with discrete data or when using nonparametric decision functions. For a detailed discussion on this matter, see Peña and Habiger (2009) and Habiger and Peña (2009).

Decision functions typically depend on a size parameter $\alpha \in [0,1]$. When viewed as a process in α, we obtain the notion of a (nonrandomized) decision process, which is a stochastic process $\Delta = \{\delta(\alpha) : \alpha \in [0,1]\}$ where, $\forall \alpha \in [0,1]$, $\delta(\alpha)$ is a decision function. We further assume the following conditions:

(D1) $\delta(0) = 0$ and $\delta(1) = 1$ a.e.-$\mathcal{P}$.

(D2) The sample paths $\alpha \mapsto \delta(\alpha)$ are, a.e.-$\mathcal{P}$, $\{0,1\}$-valued step-functions which are nondecreasing and right-continuous, that is, are *cadlag* sample paths.

Let $\mathcal{M}$ be a finite set with $|\mathcal{M}| = M$. An $\mathcal{M}$-indexed multiple decision problem is one whose action space is $\mathfrak{A}^M$. In the context of a multiple hypotheses testing problem, 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) for such a problem is a $\delta = (\delta_m : m \in \mathcal{M})$ where δ_m is a decision function, so that $\delta : (\mathcal{X}, \mathcal{F}) \to (\mathfrak{A}^M, \sigma(\mathfrak{A}^M))$. A multiple decision process (MDP) is a $\mathbf{\Delta} = (\Delta_m : m \in \mathcal{M})$ where $\Delta_m = \{\delta_m(\alpha) : \alpha \in [0,1]\}$ is a decision process.

For each $\mathbf{P} \in \mathcal{P}$, let there be subsets $\mathcal{M}_0(\mathbf{P})$ and $\mathcal{M}_1(\mathbf{P})$ of $\mathcal{M}$ such that

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

Furthermore, we assume that the following conditions hold:

(D3) Under $\mathbf{P}$, the subcollections $\{\Delta_m : m \in \mathcal{M}_0(\mathbf{P})\}$ 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 the multiple hypotheses testing situation, 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 of each other, under $\mathbf{P}$.

(D4) With $E_{\mathbf{P}}(\cdot)$ denoting the expectation operator under $\mathbf{P}$,

$$\forall \mathbf{P} \in \mathcal{P}, \forall m \in \mathcal{M}_0(\mathbf{P}), \forall \alpha \in [0,1] : E_{\mathbf{P}}\left\{\delta_m(\alpha)\right\} = \alpha. \tag{1}$$

The collection of all $\mathcal{M}$-indexed multiple decision processes satisfying conditions (D1)–(D4) will be denoted by $\mathfrak{D}$.

Let $\mathbf{A} = (A_m : m \in \mathcal{M})$ be an $\mathcal{M}$-indexed collection of measurable functions with $A_m : ([0,1], \sigma[0,1]) \to ([0,1], \sigma[0,1])$. We assume that, for each $m \in \mathcal{M}$, the following conditions are satisfied:

(A1) $A_m(0) = 0$ and $A_m(1) = 1$.

(A2) The mapping $\alpha \mapsto A_m(\alpha)$ is continuous and strictly increasing.

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

(A4) $\forall \alpha \in [0,1], \forall \mathbf{P} \in \mathcal{P}$,

$$|\mathcal{M}_0(\mathbf{P})| \max_{m \in \mathcal{M}_0(\mathbf{P})} A_m(\alpha) \leq \sum_{m \in \mathcal{M}} A_m(\alpha). \tag{2}$$

Such an $\mathbf{A}$ will be called a multiple decision size process. The collection of all $\mathcal{M}$-indexed multiple decision size processes will be denoted by $\mathfrak{S}$. A particular element of $\mathfrak{S}$ is the Sidak multiple decision size process (cf., Šidák (1967)) $\mathbf{A}^S = (A_m^S : m \in \mathcal{M})$ with

$$A_m^S(\alpha) = 1 - (1-\alpha)^{1/M},\ \alpha \in [0,1], m \in \mathcal{M}. \tag{3}$$

Before proceeding, we also introduce the notion of generalized P-value statistics. For $m \in \mathcal{M}$, let us define

$$\alpha_m \equiv \alpha_m(\mathbf{\Delta}, \mathbf{A}) = \inf\left\{\alpha \in [0,1] : \delta_m(A_m(\alpha)) = 1\right\}. \tag{4}$$

The collection $(\alpha_m(\mathbf{\Delta}, \mathbf{A}) : m \in \mathcal{M})$ could be viewed as a vector of generalized P-value statistics associated with the pair $(\mathbf{\Delta}, \mathbf{A})$. 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_ms. We shall assume without much loss of generality that these generalized P-values are a.e. $[\mathcal{P}]$ distinct, and then denote by

$$0 \equiv \alpha_{(0)} < \alpha_{(1)} < \alpha_{(2)} < \ldots < \alpha_{(M)} < 1 \equiv \alpha_{(M+1)} \tag{5}$$

the ordered generalized P-value statistics. The vector of statistics given by

$$((1), (2), \ldots, (M)), \tag{6}$$

which takes values in the space of permutations of $(1, 2, \ldots, M)$, is the vector of antiranks of the generalized P-value statistics. Thus, in the sequel, (m) denotes the index of the mth smallest P-value statistic.

2 Classes of MDFs

For a $\mathbf{\Delta} = \{\Delta_m : m \in \mathcal{M}\} \in \mathfrak{D}$, an $\mathbf{A} = \{A_m : m \in \mathcal{M}\} \in \mathfrak{S}$, a $\mathbf{P} \in \mathcal{P}$, and an $\alpha \in [0, 1]$, define

$$S_0(\alpha) \equiv S_0(\alpha; \mathbf{\Delta}, \mathbf{A}, \mathbf{P}) = \sum_{m \in \mathcal{M}_0(\mathbf{P})} \delta_m(A_m(\alpha)); \tag{7}$$

$$S(\alpha) \equiv S(\alpha; \mathbf{\Delta}, \mathbf{A}) = \sum_{m \in \mathcal{M}} \delta_m(A_m(\alpha)); \tag{8}$$

$$F(\alpha) \equiv F(\alpha; \mathbf{\Delta}, \mathbf{A}, \mathbf{P}) = \frac{S_0(\alpha)}{S(\alpha)} I\{S(\alpha) > 0\}, \tag{9}$$

with the convention that $0/0 = 0$. These quantities have the following interpretations: When δ_m has size $A_m(\alpha)$ for each $m \in \mathcal{M}$, $S_0(\alpha)$ represents the number of false discoveries, $S(\alpha)$ is the number of discoveries, and $F(\alpha)$ is the proportion of false discoveries among all discoveries.

Furthermore, for $q \in [0, 1]$, define the random variables

$$\alpha^\dagger(q) \equiv \alpha^\dagger(q; \mathbf{\Delta}, \mathbf{A}) = \sup\left\{\alpha \in [0, \alpha_{(M)}] : \prod_{m \in \mathcal{M}} [1 - A_m(\alpha)]^{1 - \delta_m(A_m(\alpha)-)} \geq 1 - q\right\}; \tag{10}$$

and

$$\alpha^*(q) \equiv \alpha^*(q;\mathbf{\Delta},\mathbf{A}) = \sup\left\{\alpha \in [0,1] : \sum_{m\in\mathcal{M}} A_m(\alpha) \le qS(\alpha;\mathbf{\Delta},\mathbf{A})\right\}. \tag{11}$$

The two major results of this paper are contained in Theorem 1 and Theorem 2. We present the statements of these theorems, but defer their proofs to Section 3.

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

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

Observe that $E_{\mathbf{P}}\left\{I\{S_0(\alpha^\dagger(q;\mathbf{\Delta},\mathbf{A});\mathbf{\Delta},\mathbf{A},\mathbf{P}) \ge 1\}\right\}$ represents the probability of committing at least one false discovery when the true underlying probability measure is $\mathbf{P}$. Thus, Theorem 1 shows that for any $q \in [0,1]$, any multiple decision process $\mathbf{\Delta} \in \mathfrak{D}$, and multiple decision size process $\mathbf{A} \in \mathfrak{S}$, the MDF defined via

$$\delta^\dagger(q) \equiv \delta^\dagger(q;\mathbf{\Delta},\mathbf{A}) = \left(\delta_m[A_m(\alpha^\dagger(q;\mathbf{\Delta},\mathbf{A}))]:\ m \in \mathcal{M}\right), \tag{12}$$

strongly controls the family-wise error rate (FWER) at q.

Theorem 2 *Under conditions (D1)–(D4) for $\mathfrak{D}$ and (A1)–(A4) for $\mathfrak{S}$,*

$$\forall \mathbf{P} \in \mathcal{P}, \forall \mathbf{\Delta} \in \mathfrak{D}, \forall \mathbf{A} \in \mathfrak{S}, \forall q \in [0,1]:\ E_{\mathbf{P}}\left\{F(\alpha^*(q;\mathbf{\Delta},\mathbf{A});\mathbf{\Delta},\mathbf{A},\mathbf{P})\right\} \le q.$$

The quantity $E_{\mathbf{P}}\left\{F(\alpha^*(q;\mathbf{\Delta},\mathbf{A});\mathbf{\Delta},\mathbf{A},\mathbf{P})\right\}$ is called the false discovery rate (FDR) as introduced in the seminal paper of Benjamini and Hochberg (1995). The implication of Theorem 2 is that if, for each $q \in [0,1]$, and for any multiple decision process $\mathbf{\Delta} \in \mathfrak{D}$ and multiple decision size process $\mathbf{A} \in \mathfrak{S}$, we define the multiple decision function

$$\delta^*(q) \equiv \delta^*(q;\mathbf{\Delta},\mathbf{A}) = \left(\delta_m[A_m(\alpha^*(q;\mathbf{\Delta},\mathbf{A}))]:\ m \in \mathcal{M}\right), \tag{13}$$

then $\delta^*(q)$ is a multiple decision function which strongly controls the FDR at q. We point out that there are other types of false discovery rates such as the marginal false discovery rate (cf., Efron, Tibshirani, Storey, and Tusher (2001), Efron (2004), Efron (2007), and Sun and Cai (2007)), or the positive false discovery rate (cf., Storey (2003)). We do not consider these types of false discovery rates in this paper since, as pointed out in Benjamini and Hochberg (1995), they cannot be controlled strongly. See Genovese and Wasserman (2002) and Storey (2003) for a comparison of these varied types of false discovery rates.

The importance of the preceding results is that each multiple decision process $\mathbf{\Delta} \in \mathfrak{D}$ may have an associated multiple decision size process $\mathbf{A} \equiv \mathbf{A}(\mathbf{\Delta}) \in \mathfrak{S}$ 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. To define this rate, let

$$M(\alpha) \equiv M(\alpha; \mathbf{\Delta}, \mathbf{A}, \mathbf{P}) = \frac{\sum_{m \in \mathcal{M}_1(\mathbf{P})} (1 - \delta_m(A_m(\alpha)))}{|\mathcal{M}_1(\mathbf{P})|} I\{|\mathcal{M}_1(\mathbf{P})| > 0\}. \tag{14}$$

The quantity $M(\alpha)$ has the interpretation of being the proportion of missed discoveries relative to the number of correct alternative hypotheses. The missed discovery rate (MDR) of the multiple decision function in (13) is then $E_{\mathbf{P}} \left\{ M(\alpha^*(q); \mathbf{\Delta}, \mathbf{A}, \mathbf{P}) \right\}$. For the given $\mathbf{\Delta}$, with proper choice of $\mathbf{A}$, we may be able to find a multiple decision function that strongly controls the FWER or the FDR, while at the same time possessing an optimal property with respect to another criterion, such as having a small, maximally over $\mathcal{P}$, MDR. See an implementation of this idea in Peña and Habiger (2009) in the situation where each pair of hypotheses deals with a simple null hypothesis versus simple alternative hypothesis.

Previous works, such as the paper of Benjamini and Hochberg (1995) which introduced the notion of the FDR and an FDR-controlling procedure, focussed on developing a multiple decision function and then verifying that it controls the FDR. By providing a class of MDFs where each member (strongly) controls the FWER, given by

$$\mathfrak{C}^\dagger = \left\{ \delta^\dagger(q; \mathbf{\Delta}, \mathbf{A}) : \ \mathbf{\Delta} \in \mathfrak{D}, \mathbf{A} \in \mathfrak{S} \right\}; \tag{15}$$

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

$$\mathfrak{C}^* = \left\{ \delta^*(q; \mathbf{\Delta}, \mathbf{A}) : \ \mathbf{\Delta} \in \mathfrak{D}, \mathbf{A} \in \mathfrak{S} \right\}, \tag{16}$$

then we acquire the possibility of selecting from these classes a multiple decision function which possesses some other desirable property. More discussion of this issue will be provided in Section 5.

3 Proofs

3.1 Of Theorem 1

Let $\mathbf{\Delta} \in \mathfrak{D}$ and $\mathbf{A} \in \mathfrak{S}$ be fixed decision process and size process, respectively. Let $q \in [0,1]$ and denote by $\mathbf{P} \in \mathcal{P}$ the (unknown) underlying probability measure. We suppress writing the arguments $\mathbf{\Delta}$, $\mathbf{A}$, and $\mathbf{P}$ in S_0, $\alpha^\dagger$, and the soon-to-be-introduced $\alpha^\#$. Let us first define the stochastic process $\{H_1(\alpha) : \alpha \in [0,1]\}$ via

$$H_1(\alpha) \equiv H_1(\alpha; \mathbf{\Delta}, \mathbf{A}) = \prod_{m \in \mathcal{M}} [1 - A_m(\alpha)]^{1-\delta_m(A_m(\alpha)-)}. \tag{17}$$

The sample paths of this process are, a.e. [$\mathbf{P}$], left-continuous with right-hand limits (*caglad*) and are piecewise nonincreasing with $H_1(\alpha-) = H_1(\alpha) \leq H_1(\alpha+)$ for every $\alpha \in [0,1]$. In terms of H_1, we have that

$$\alpha^\dagger(q) = \sup\left\{\alpha \in [0, \alpha_{(M)}] : H_1(\alpha) \geq 1 - q\right\}.$$

Observe that we always have

$$H_1(\alpha^\dagger(q)) \geq 1 - q. \tag{18}$$

Now, considering the quantity of main interest to us, we have

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

Observe that

$$\bigcap_{m \in \mathcal{M}_0(\mathbf{P})} \left[\delta_m(A_m(\alpha^\dagger(q))) = 0\right] = \left\{\alpha^\dagger(q) < \min_{m \in \mathcal{M}_0(\mathbf{P})} \alpha_m\right\}. \tag{19}$$

Next, let us define the stochastic process $\{H_2(\alpha) : \alpha \in [0,1]\}$ via

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

Analogously to the H_1 function, this has caglad sample paths. Let us then define the quantity

$$\alpha^{\#}(q) \equiv \alpha^{\#}(q; \mathbf{\Delta}, \mathbf{A}, \mathbf{P}) = \sup\left\{\alpha \in [0, \alpha_{(M)}] : H_2(\alpha) \geq 1 - q\right\}.$$

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, we also note that

$$H_2(\alpha^{\#}(q)) \geq 1 - q. \tag{20}$$

The importance of this quantity $\alpha^{\#}(q)$ arises by observing that

$$\left\{\alpha^{\dagger}(q) < \min_{m \in \mathcal{M}_0(\mathbf{P})} \alpha_m\right\} = \left\{\alpha^{\#}(q) < \min_{m \in \mathcal{M}_0(\mathbf{P})} \alpha_m\right\}. \tag{21}$$

Therefore, by (19), (21), and the iterated expectation rule,

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

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),

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

with the last equality following from condition (D4). It follows that

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

with the last inequality following from (20). Thus, finally, we have

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

This completes the proof of Theorem 1.

3.2 Of Theorem 2

The proof of this theorem builds on a proof presented in Peña and Habiger (2009) for a more restrictive setting. Indeed, the idea of providing a *class* of FDR-controlling multiple decision functions arose as a consequence of the aforementioned manuscript upon observing that the proof of one of the main results there (Theorem 9.1) is actually independent of the choice of the size functions.

Let us fix a $\mathbf{\Delta}\in\mathfrak{D}$ and an $\mathbf{A}\in\mathfrak{S}$, and let $q\in(0,1]$. Below, we will suppress writing $\mathbf{\Delta}$ and $\mathbf{A}$ in the processes and random variables, unless necessary, as when we want to use in conjunction with the Sidak sizes.

First, observe that the case $q=0$ is trivial since $\alpha^*(0)=0$ so that $F(\alpha^*(0))=0$, thus the restriction to $q\in(0,1]$. By the defining property of $\alpha^*(q)$ given in (11), we have that

$$S(\alpha^*(q)) \geq \frac{1}{q}A_\bullet(\alpha^*(q)) \tag{22}$$

where $A_\bullet(\alpha) = \sum_{m\in\mathcal{M}} A_m(\alpha)$. Consequently from (9),

$$F(\alpha^*(q)) \le q\frac{S_0(\alpha^*(q))}{A_\bullet(\alpha^*(q))} I\{S(\alpha^*(q) > 0\} \le q\frac{S_0(\alpha^*(q))}{A_\bullet(\alpha^*(q))}. \tag{23}$$

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

$$\mathcal{F}_\alpha \equiv \mathcal{F}_\alpha(\mathbf{\Delta}, \mathbf{A}) = \sigma\left\{\delta_m(A_m(\beta)) :\ \beta \in [\alpha, 1], m \in \mathcal{M}\right\}. \tag{24}$$

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 $(T_0(\alpha) :\ \alpha \in [0,1])$ 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 $0 \le \alpha \le \beta \le 1$ and let $\mathbf{P} \in \mathcal{P}$. Then, since $\delta_m \in \{0,1\}$, we have

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

The second equality follows from (D3), whereas the second to last equality follows since

$$\begin{aligned}
E_{\mathbf{P}}\left\{\frac{\delta_m(A_m(\alpha))}{A_m(\alpha)}|\delta_m(A_m(\beta)) = 1\right\} &= \frac{\mathbf{P}\{\delta_m(A_m(\alpha)) = 1, \delta_m(A_m(\beta)) = 1\}}{\mathbf{P}\{\delta_m(A_m(\beta)) = 1\}} \\
&= \frac{\mathbf{P}\{\delta_m(A_m(\alpha)) = 1\}}{\mathbf{P}\{\delta_m(A_m(\beta)) = 1\}} \\
&= \frac{A_m(\alpha)}{A_m(\beta)}
\end{aligned}$$

because of condition (A2) for the $A_m(\cdot)$s and condition (D2) for the $\delta_m(\cdot)$s, and (1) in condition (D4). 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, $E_{\mathbf{P}}(T_0(1)) = |\mathcal{M}_0(\mathbf{P})|$.

Now, from the inequality in (23), we obtain

$$\begin{aligned} E_{\mathbf{P}}[F(\alpha^*(q))] &\leq qE_{\mathbf{P}}\left[\frac{S_0(\alpha^*(q))}{A_\bullet(\alpha^*(q))}\right] \\ &= q\sum_{m\in\mathcal{M}_0(\mathbf{P})} E_{\mathbf{P}}\left[\frac{\delta_m(\alpha^*(q))}{A_\bullet(\alpha^*(q))}\right] \\ &= q\sum_{m\in\mathcal{M}_0(\mathbf{P})} E_{\mathbf{P}}\left[\frac{\delta_m(\alpha^*(q))}{A_m(\alpha^*(q))}\frac{A_m(\alpha^*(q))}{A_\bullet(\alpha^*(q))}\right] \\ &\leq q\left[\sup_{\alpha\in[0,1]}\max_{m\in\mathcal{M}_0(\mathbf{P})}\frac{A_m(\alpha)}{A_\bullet(\alpha)}\right]E_{\mathbf{P}}\left[T_0(\alpha^*(q))\right] \\ &\leq q\frac{1}{|\mathcal{M}_0(\mathbf{P})|}E_{\mathbf{P}}[T_0(1)] \\ &= q\frac{|\mathcal{M}_0(\mathbf{P})|}{|\mathcal{M}_0(\mathbf{P})|} \\ &= q, \end{aligned}$$

where the last inequality is obtained using condition (A4) and by invoking the Optional Sampling Theorem for (reverse) martingales (cf., Doob (1953)), and the second to last equality because of $E_{\mathbf{P}}[T_0(1)] = |\mathcal{M}_0(\mathbf{P})|$.

Note that, in particular, since the Sidak multiple decision size process $\mathbf{A}^S$ always satisfies condition (A4) for *all* $\mathbf{P} \in \mathcal{P}$, then $\forall\mathbf{\Delta} \in \mathfrak{D}, \forall\mathbf{P} \in \mathcal{P}$, we have the property

$$E_{\mathbf{P}}\left\{F(\alpha^*(q;\mathbf{\Delta},\mathbf{A}^S);\mathbf{\Delta},\mathbf{A}^S)\right\} \leq q. \tag{25}$$

Let us denote by $\mathcal{P}_0 = \{\mathbf{P} \in \mathcal{P} : \ \mathcal{M}_0(\mathbf{P}) = \mathcal{M}\}$. Observe that for $\mathbf{P} \in \mathcal{P}_0$, condition (A4) will not be satisfied unless $\mathbf{A} = \mathbf{A}^S$. We still therefore need to establish that for an arbitrary $\mathbf{A} \in \mathfrak{S}$ and a $\mathbf{P} \in \mathcal{P}_0$,

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

For such a $\mathbf{P} \in \mathcal{P}_0$, we have $F(\alpha;\mathbf{\Delta},\mathbf{A}) = I\{S(\alpha;\mathbf{\Delta},\mathbf{A}) > 0\}$, so that

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

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

$$\{\alpha^*(q;\mathbf{\Delta},\mathbf{A}) > 0\} = \bigcup_{\alpha\in(0,1]}\left\{\frac{S(\alpha;\mathbf{\Delta},\mathbf{A})}{A_\bullet(\alpha)} \geq \frac{1}{q}\right\}. \tag{26}$$

In Lemma 9.1 of Peña and Habiger (2009) it was established, using an inequality of Hoeffding (1956), that for $W_m(\eta_m), m \in \mathcal{M}$, independent Bernoulli($\eta_m$) random variables with $\eta_m \in [0,1]$ and satisfying $\prod_{m\in\mathcal{M}}(1-\eta_m) = 1-\alpha$, for each $t \geq 1$,

$$\mathbf{P}\left\{\frac{\sum_{m\in\mathcal{M}} W_m(\eta_m)}{\sum_{m\in\mathcal{M}} \eta_m} \geq t\right\} \leq \mathbf{P}\left\{\frac{\sum_{m\in\mathcal{M}} W_m(\eta_m^S)}{\sum_{m\in\mathcal{M}} \eta_m^S} \geq t\right\}, \tag{27}$$

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

By noting that, under $\mathbf{P} \in \mathcal{P}_0$, the $\delta_m(A_m(\alpha))$s are independent Bernoulli($A_m(\alpha)$), then by using the inequality in (27) and condition (A3), it follows for $q \in (0,1]$ that

$$\mathbf{P}\left\{\frac{S(\alpha;\mathbf{\Delta},\mathbf{A})}{A_\bullet(\alpha)} \geq \frac{1}{q}\right\} \leq \mathbf{P}\left\{\frac{S(\alpha;\mathbf{\Delta},\mathbf{A}^S)}{A_\bullet^S(\alpha)} \geq \frac{1}{q}\right\}. \tag{28}$$

Combining the results in (26) and (28), we obtain

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

But since we have already established that for $\mathbf{P} \in \mathcal{P}_0$, $\mathbf{P}\{\alpha^*(q;\mathbf{\Delta},\mathbf{A}^S) > 0\} \leq q$, then $\mathbf{P}\{\alpha^*(q;\mathbf{\Delta},\mathbf{A}) > 0\} \leq q$, so $E_{\mathbf{P}}\{F(\alpha^*(q;\mathbf{\Delta},\mathbf{A});\mathbf{\Delta},\mathbf{A})\} \leq q$ for any $\mathbf{P} \in \mathcal{P}_0$.

This completes the proof of Theorem 2.

We remark that the martingale-based proof above provides an alternative proof of the Benjamini and Hochberg (1995) proof of FDR control by their procedure since, as will be shown in Section 5, their procedure is a special case of δ^* when the size process is taken to be the Sidak size process. A martingale proof of this FDR control was also actually presented in Storey, Taylor, and Siegmund (2004), but specifically only for the BH procedure.

4 In Terms of the Generalized P-Value Statistics

In this section we provide equivalent formulations in terms of the generalized P-value statistics $(\alpha_m, m \in \mathcal{M})$ defined in (4). Recalling that $\alpha_{(m)}$s denote the ordered generalized P-value statistics, note that for $\alpha \in [\alpha_{(k)}, \alpha_{(k+1)})$ with $k \in \{0,1,2,\ldots,M\}$, we have $S(\alpha) = \sum_{m=1}^M \delta_m(A_m(\alpha)) = k$, or $\delta_{(m)}(A_{(m)}(\alpha)) = 1$ for $m = 1,2,\ldots,k$, and $\delta_{(m)}(A_{(m)}(\alpha)) = 0$ for $m = k+1, k+2, \ldots, M$.

For the specified $q \in [0,1]$, $\mathbf{\Delta}$, and $\mathbf{A}$ in the definitions of $\alpha^\dagger(q)$ and $\alpha^*(q)$ in (10) and (11), respectively, define the random variables

$$\begin{aligned} J^\dagger(q) &= \max\left\{m \in \{0,1,2,\ldots,M\} : \prod_{j=1}^{M}[1 - A_j(\alpha_{(m)})]^{1-\delta_j(A_j(\alpha_{(m)})-)} \geq 1-q\right\} \\ &= \max\left\{m \in \{0,1,2,\ldots,M\} : \prod_{j=m}^{M}[1 - A_{(j)}(\alpha_{(m)})] \geq 1-q\right\} \end{aligned} \tag{29}$$

and

$$\begin{aligned} J^*(q) &= \max\left\{m \in \{0,1,2,\ldots,M\} : \sum_{j=1}^{M} A_j(\alpha_{(m)}) \leq qS(\alpha_{(m)})\right\} \\ &= \max\left\{m \in \{0,1,2,\ldots,M\} : \sum_{j=1}^{M} A_{(j)}(\alpha_{(m)}) \leq qm\right\}. \end{aligned} \tag{30}$$

Then, it is easy to see from the definitions of $\alpha^\dagger(q)$, $J^\dagger(q)$, $\alpha^*(q)$, $J^*(q)$, and the generalized P-value statistics, together with the observation that the sample paths $\alpha \mapsto H_1(\alpha)$ (see (17)) are caglad and decreasing on each of the intervals $[\alpha_{(m)}, \alpha_{(m+1)}), m = 0,1,2\ldots,M$, that

$$\alpha^\dagger(q) \in \left[\alpha_{(J^\dagger(q))}, \alpha_{(J^\dagger(q)+1)}\right).$$

Also, it is easy to see that

$$\alpha^*(q) \in \left[\alpha_{(J^*(q))}, \alpha_{(J^*(q)+1)}\right).$$

As a consequence, the $\delta^\dagger(q)$ in Theorem 1 and the $\delta^*(q)$ in Theorem 2 could be re-expressed in terms of the ordered generalized P-value statistics, $J^\dagger(q)$, and $J^*(q)$, via

$$\delta^\dagger(q) \equiv \left(\delta_m(A_m(\alpha^\dagger(q))), m \in \mathcal{M}\right) = \left(\delta_m(A_m(\alpha_{(J^\dagger(q))})), m \in \mathcal{M}\right) \tag{31}$$

and

$$\delta^*(q) \equiv (\delta_m(A_m(\alpha^*(q))), m \in \mathcal{M}) = \left(\delta_m(A_m(\alpha_{(J^*(q))})), m \in \mathcal{M}\right). \tag{32}$$

These equivalent representations provide alternative computational approaches since, instead of computing $\alpha^\dagger(q)$ and $\alpha^*(q)$, we may just compute the generalized P-values and then the values of $J^\dagger(q)$ and $J^*(q)$.

5 Some Applications

We demonstrate in this section some applications of Theorems 1 and 2. In particular, we show that by taking the size process to be of the Sidak type in (3), the sequential Sidak FWER-controlling procedure is a special case of $\delta^\dagger(q; \mathbf{\Delta}, \mathbf{A})$ in Theorem 1; whereas, the BH FDR-controlling procedure (Benjamini and Hochberg (1995)) is a special case of $\delta^*(q; \mathbf{\Delta}, \mathbf{A})$ in Theorem 2.

5.1 Sidak Size Process

Let $\mathbf{\Delta}$ be a fixed decision process and let $q \in [0,1]$ be a pre-specified FWER or FDR level. Choose the size process to be $\mathbf{A} = \mathbf{A}^S$, so that $A_m(\alpha) = 1 - (1-\alpha)^{1/M}, m \in \mathcal{M}$. From (29), we then have

$$J^\dagger(q) = \max\left\{m \in \{0,1,2,\ldots,M\}: \ \alpha_{(m)} \leq 1 - (1-q)^{M/(M-m+1)}\right\}. \tag{33}$$

The MDF then has $\delta_{(m)} = 1$ if and only if $m \leq J^\dagger(q)$, or equivalently, $\delta_m = 1$ if and only if $\alpha_m \leq \alpha_{(J^\dagger(q))}$.

Let us relate this to the usual P-value statistics of the δ_ms. As pointed out after the definition of the generalized P-value statistics in (4), the usual P-value statistic for δ_m is given by $P_m = A_m(\alpha_m)$ which for the Sidak size process becomes

$$P_m = 1 - (1-\alpha_m)^{1/M}, m \in \mathcal{M}. \tag{34}$$

Consequently, it follows from (33) that $J^\dagger(q)$ can be expressed in terms of the ordered P_ms via

$$J^\dagger(q) = \max\left\{m \in \{0,1,2,\ldots,M\}: \ P_{(m)} \leq 1 - (1-q)^{1/(M-m+1)}\right\}. \tag{35}$$

The resulting cut-off with respect to the usual P-values would then be $P_{(J^\dagger(q))}$, so that the resulting MDF would have, for each $m \in \mathcal{M}$, $\delta_{(m)} = 1$ if and only if $m \leq J^\dagger(q)$. Equivalently, $\delta_m = 1$ if and only if $P_m \leq P_{(J^\dagger(q))}$. This procedure is the sequential Sidak procedure (cf., Dudoit, Shaffer, and Boldrick (2003), Dudoit and van der Laan (2008)) which strongly controls the FWER at level q. Thus, the sequential Sidak procedure arises from Theorem 1 by choosing the Sidak size process for whatever choice of the decision process $\mathbf{\Delta}$.

Next, consider the same setting but now using Theorem 2. From (30), we have for the Sidak size process,

$$\begin{aligned} J^*(q) &= \max\left\{m \in \{0,1,2,\ldots,M\}:\ M\left[1-(1-\alpha_{(m)})^{1/M}\right] \leq qm\right\} \\ &= \max\left\{m \in \{0,1,2,\ldots,M\}:\ P_{(m)} \leq \frac{qm}{M}\right\}. \end{aligned} \tag{36}$$

Thus, the resulting procedure has, for $m \in \mathcal{M}$, $\delta_{(m)} = 1$ if and only if $m \leq J^*(q)$. Equivalently, $\delta_m = 1$ if and only if $P_m \leq P_{(J^*(q))}$. This is precisely the BH procedure which controls the FDR at level q. Thus, it arises from Theorem 2 by choosing the Sidak size for whatever choice of the decision process $\mathbf{\Delta}$.

5.2 Towards Optimal MDFs

Finally, in this subsection, we indicate, without going into much detail, 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 so-called non-exchangeable multiple hypotheses testing settings. Let us fix a decision process $\mathbf{\Delta} \in \mathfrak{D}$ and fix a probability measure $\mathbf{P}_1 \in \mathcal{P}$. Define the mappings $\pi_m : [0,1] \to [0,1]$ for $m \in \mathcal{M}$ according to

$$\pi_m(\alpha; \mathbf{P}_1) = E_{\mathbf{P}_1}[\delta_m(\alpha)],\ \alpha \in [0,1]. \tag{37}$$

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 $\mathbf{P}_1$, and in this case $\pi_m(\cdot; \mathbf{P}_1)$ is the receiver operating characteristic curve (ROC). Assume that for each $m \in \mathcal{M}$, $\alpha \mapsto \pi_m(\alpha; \mathbf{P}_1)$ is strictly increasing with $\pi_m(1; \mathbf{P}_1) = 1$ and twice-differentiable.

Suppose that it is desired to strongly control the overall FWER or FDR at some prespecified level $q \in [0,1]$, and at the same time maximize the total (or average) power at $\mathbf{P} = \mathbf{P}_1$. Our idea, partly implemented in Peña and Habiger (2009), is to first obtain the optimal size process for *weak* FWER control associated with $\mathbf{\Delta}$, denoted by $\mathbf{A}^* = (A_m^*(\alpha), m \in \mathcal{M})$, which is the size process $\mathbf{A}$ satisfying the condition

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

and such that the total power, at $\mathbf{P} = \mathbf{P}_1$, given by $\sum_{m=1}^{M} \pi_m(A_m(\alpha); \mathbf{P}_1)$, is maximized. Under regularity conditions on the ROC functions, the optimal $\mathbf{A}^*$ process could be obtained using Lagrange optimization, cf., Theorem 5.1 in Peña and Habiger (2009).

Now, having determined the optimal size process $\mathbf{A}^*$ associated with $\mathbf{\Delta}$, which is at this point optimal only in the sense of weak FWER control, we then apply Theorem 1 to obtain the MDF $\delta^\dagger(q; \mathbf{D}, \mathbf{A}^*)$ which will strongly control the FWER at q; or apply Theorem 2 to obtain the MDF $\delta^*(q; \mathbf{D}, \mathbf{A}^*)$ which will control the FDR at q. By virtue of the choice of the size process $\mathbf{A}^*$, which is tied-in to the decision process $\mathbf{D}$ and the target probability measure $\mathbf{P}_1$, we expect that the MDFs $\delta^\dagger(q; \mathbf{D}, \mathbf{A}^*)$ and $\delta^*(q; \mathbf{D}, \mathbf{A}^*)$ will perform better with respect to overall power at $\mathbf{P}_1$ relative to, for example, the sequential Sidak or the BH procedures which arise from non-optimal size processes. See, for example, the results of a modest simulation study in Peña and Habiger (2009) demonstrating the improvement over the BH procedure of the MDF δ^* in a more specific setting. Further improvements could be obtained by a proper choice of the decision process $\mathbf{\Delta}$, such as for instance choosing it to have components that are uniformly most powerful (UMP) or uniformly most powerful unbiased (UMPU) test functions. These issues will be further considered in future work.

We also mention that there are other approaches to searching for multiple decision functions with optimality properties such as those in Westfall, Krishen, and Young (1998), Westfall and Krishen (2001), Genovese and Wasserman (2003), Genovese, Roeder, and Wasserman (2006), Storey (2007), Sun and Cai (2007), and Sarkar, Zhou, and Ghosh (2008), or via the Bayesian or empirical Bayes paradigms such as in Scott and Berger (2006) and Efron (2008). Extensions of Theorem 1 and Theorem 2 for settings where $\{\delta_m : m \in \mathcal{M}_0(\mathbf{P})\}$ are dependent may also be possible. Such results may serve as generalizations of those in Sarkar and

Chang (1997) and Benjamini and Yekutieli (2001) which deal with dependencies; see also the review article by Sarkar (2008).

Acknowledgements

Some ideas and results in this paper were obtained while the first author was a Faculty Fellow at the Statistical and Applied Mathematical Sciences Institute (SAMSI) during his sabbatical leave from the University of South Carolina in Fall 2008. He is grateful to SAMSI Director Dr. James Berger for facilitating his stay at the Institute. As such, this work was partially supported by the National Science Foundation (NSF) under Grant DMS-0635449 to SAMSI. However, any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. Also, some of the main ideas and results on this paper were obtained while the first author was visiting the University of Oslo in Oslo, Norway and the Norwegian University of Science and Technology (NTNU) in Trondheim, Norway in Fall 2008. He wishes to thank Professor Odd Aalen and Professor Bo Lindqvist for facilitating his visit to these universities. The authors wish to acknowledge the following grants which partially supported this research: NSF Grant DMS 0805809, National Institutes of Health (NIH) Grant RR17698, and Environmental Protection Agency (EPA) Grant RD-83241902-0 to University of Arizona with subaward number Y481344 to the University of South Carolina.

References

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

Benjamini, Y. and Yekutieli, D. (2001), "The control of the false discovery rate in multiple testing under dependency," *Ann. Statist.*, 29, 1165–1188.

Doob, J. L. (1953), *Stochastic processes*, New York: John Wiley & Sons Inc.

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

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

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

— (2007), "Size, power and false discovery rates," *Ann. Statist.*, 35, 1351–1377.

— (2008), "Microarrays, empirical Bayes and the two-groups model," *Statist. Sci.*, 23, 1–22.

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

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

— (2003), "Bayesian and frequentist multiple testing," in *Bayesian statistics, 7 (Tenerife, 2002)*, New York: Oxford Univ. Press, pp. 145–161, with discussions by Merlise A. Clyde and Christian P. Robert and Judith Rousseau, and a reply by the authors.

Genovese, C. R., Roeder, K., and Wasserman, L. (2006), "False discovery control with p-value weighting," *Biometrika*, 93, 509–524.

Habiger, J. and Peña, E. A. (2009), "Randomized P-Values in Multiple Testing," Tech. rep., University of South Carolina.

Hoeffding, W. (1956), "On the distribution of the number of successes in independent trials," *Ann. Math. Statist.*, 27, 713–721.

Peña, E. A. and Habiger, J. (2009), "Power-Enhanced Multiple Decision Functions Controlling Family-Wise Error and False Discovery Rates," Tech. rep., University of South Carolina, Columbia, South Carolina.

Sarkar, S. K. (2008), "On Methods Controlling the False Discovery Rate (with discussion)," *Sankhya*, In Press.

Sarkar, S. K. and Chang, C.-K. (1997), "The Simes method for multiple hypothesis testing with positively dependent test statistics," *J. Amer. Statist. Assoc.*, 92, 1601–1608.

Sarkar, S. K., Zhou, T., and Ghosh, D. (2008), "A general decision theoretic formulation of procedures controlling FDR and FNR from a Bayesian perspective," *Statist. Sinica*, 18, 925–945.

Scott, J. and Berger, J. (2006), "An exploration of aspects of Bayesian multiple testing," *Journal of Statistical Planning and Inference*, 136, 2144–2162.

Šidák, Z. (1967), "Rectangular confidence regions for the means of multivariate normal distributions," *J. Amer. Statist. Assoc.*, 62, 626–633.

Storey, J. (2003), "The positive false discovery rate: a Bayesian interpretation and the q-Value," *The Annals of Statistics*, 31, 2012 – 2035.

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

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

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

Westfall, P. H. and Krishen, A. (2001), "Optimally weighted, fixed sequence and gatekeeper multiple testing procedures," *J. Statist. Plann. Inference*, 99, 25–40.

Westfall, P. H., Krishen, A., and Young, S. S. (1998), "Using Prior Information to Allocate Significance Levels for Multiple Endpoints," *Statistics in Medicine*, 17, 2107–2119.