University of California, Berkeley
U.C. Berkeley Division of Biostatistics Working Paper Series

Year 2004 *Paper* 141

Multiple Testing. Part III. Procedures for Control of the Generalized Family-Wise Error Rate and Proportion of False Positives

Mark J. van der Laan* Sandrine Dudoit†
Katherine S. Pollard‡

*Division of Biostatistics, School of Public Health, University of California, Berkeley, laan@berkeley.edu

†Division of Biostatistics, School of Public Health, University of California, Berkeley, sandrine@stat.berkeley.edu

‡Center for Biomolecular Science & Engineering, University of California, Santa Cruz, kpollard@gladstone.ucsf.edu

http://biostats.bepress.com/ucbbiostat/paper141

Multiple Testing. Part III. Procedures for Control of the Generalized Family-Wise Error Rate and Proportion of False Positives

Mark J. van der Laan, Sandrine Dudoit, and Katherine S. Pollard

Abstract

The accompanying articles by Dudoit et al. (2003b) and van der Laan et al. (2003) provide single-step and step-down resampling-based multiple testing procedures that asymptotically control the family-wise error rate (FWER) for general null hypotheses and test statistics. The proposed procedures fundamentally differ from existing approaches in the choice of null distribution for deriving cut-offs for the test statistics and are shown to provide asymptotic control of the FWER under general data generating distributions, without the need for conditions such as subset pivotality. In this article, we show that any multiple testing procedure (asymptotically) controlling the FWER at level alpha can be augmented into: (i) a multiple testing procedure (asymptotically) controlling the generalized family-wise error rate (i.e., the probability, gFWER(k), of having more than k false positives) at level alpha and (ii) a multiple testing procedure (asymptotically) controlling the probability, PFP(q), that the proportion of false positives among the rejected hypotheses exceeds a user-supplied value q in (0,1) at level alpha. Existing procedures for control of the proportion of false positives typically rely on the assumption that the test statistics are independent, while our proposed augmentation procedures control the PFP and gFWER for general data generating distributions, with arbitrary dependence structures among variables. Applying our augmentation methods to step-down multiple testing procedures that asymptotically control the FWER at exact level alpha (van der Laan et al., 2003), yields multiple testing procedures that also asymptotically control the gFWER and PFP at exact level alpha. Finally, the adjusted p-values for the gFWER and PFP-controlling augmentation procedures are shown to be simple functions of the adjusted p-values for the original FWER-controlling procedure.

Contents

1 Introduction

Multiple hypothesis testing methods are concerned with the simultaneous test of $m > 1$ null hypotheses, while controlling a suitably defined Type I error (i.e., false positive) rate. There is a rich literature on multiple testing. We refer the reader to textbooks by Hochberg and Tamhane (1987) and Westfall and Young (1993), and to overview articles by Shaffer (1995) and Dudoit et al. (2003a). Classical approaches to multiple testing call for controlling the family-wise error rate (FWER), that is, the chance of making at least one Type I error. Hochberg and Tamhane (1987) describe a variety of single-step and step-down methods for controlling the FWER, based on cut-off rules for the ordered p-values corresponding to each null hypothesis. Westfall and Young (1993) provide resampling-based multiple testing procedures for control of the FWER. Pollard and van der Laan (2003) and Dudoit et al. (2003b) propose a general class of single-step multiple testing procedures for controlling a user-supplied Type I error rate, defined as an arbitrary parameter of the distribution of the number of false positives. Such error rates include, for example, tail probabilities for the number of Type I errors (generalized family-wise error rate, gFWER), and the mean (per comparison error rate, PCER) and median number of Type I errors. While Pollard and van der Laan (2003) focus on single-parameter null hypotheses and test statistics which are asymptotically normally distributed, Dudoit et al. (2003b) extend the proposed approach to general null hypotheses defined in terms of submodels for the data generating distribution, provide adjusted p-values for each procedure, and present a number of applications of the theory. van der Laan et al. (2003) propose step-down multiple testing procedures for controlling the FWER for general null hypotheses and test statistics. As detailed in these three articles, our approach to multiple testing fundamentally differs from existing approaches in the literature (e.g., methods discussed in Hochberg and Tamhane (1987) and Westfall and Young (1993)) in the choice of null distribution for deriving rejection regions, i.e., cut-offs for the test statistics. We propose a general characterization and explicit construction for a test statistics null distribution that provides asymptotic control of the Type I error rate under general data generating distributions, without the need for conditions such as subset pivotality.

A common criticism of multiple testing procedures designed to control the FWER is their lack of power, especially for large-scale testing problems such as those encountered in genomics and astronomy. Here, we consider two

broad classes of Type I error rates that are less stringent than the FWER and may therefore be more appropriate for current high-dimensional applications: the generalized family-wise error rate (gFWER) and the proportion of false positives among the rejected hypotheses (PFP). The gFWER is a relaxed version of the FWER, which allows $k > 0$ false positives, that is, $gFWER(k)$ is defined as the chance of at least $(k+1)$ Type I errors ($k = 0$ for the usual FWER). While Pollard and van der Laan (2003) and Dudoit et al. (2003b) provide single-step procedures for control of the gFWER, we note that none of the above articles supply step-down methods for controlling this error rate.

A variety of multiple testing methods controlling the false discovery rate (FDR), i.e., the expected proportion of false positives among the rejected hypotheses, have been proposed in a frequentist setting by Benjamini and co-authors (Benjamini and Hochberg, 1995; Benjamini and Yekutieli, 2001; Benjamini and Braun, 2002; Yekutieli and Benjamini, 1999). These articles establish FDR control results under the assumption that the test statistics are either independently distributed or have certain forms of dependence (e.g., positive regression dependence). Among the many applications of FDR-controlling procedures we mention Reiner et al. (2003) and Yekutieli and Benjamini (1999). Abramovich et al. (2000) show that FDR-controlling procedures can also be used to adapt to unknown sparsity in regression. A (Bayesian) mixture model approach to obtain multiple testing procedures controlling the FDR is considered in Efron et al. (2001a), Efron et al. (2001b), Storey (2001), Storey (2002), Storey and Tibshirani (2001), Storey and Tibshirani (2003), Storey et al. (2003), and Tusher et al. (2001). The latter framework assumes that test statistics are independent and identically distributed realizations from a mixture model, though robustness against this assumption has been noted.

In contrast to FDR-controlling procedures that focus on the *expected value* of the proportion of false positives among the rejected hypotheses, Genovese and Wasserman (2001) consider *tail probabilities* for this proportion. Under the assumption that the test statistics are independent, these authors provide procedures that control the probability, $PFP(q)$, that the proportion of false positives exceeds a user-supplied value $q \in (0, 1)$. Under the same independence assumption, Genovese and Wasserman (2002) extend the theory for FDR control by representing the FDR as an empirical process. In studies in which investigators wish to have high confidence (i.e., chance at least $1 - \alpha$) that the set of rejected null hypotheses contains at most a specified proportion q of false positives, PFP control is the appropriate form of

Type I error control, while FDR-controlling approaches would only control this proportion on average. Multiple testing procedures proposed thus far for controlling a parameter (i.e., mean for FDR or tail probability for $PFP(q)$) of the distribution of the proportion of false positives among the rejected hypotheses rely on the following assumptions concerning the joint distribution of the test statistics: independence, specific dependence structures (such as positive regression dependence), or normality.

The present article proposes simple augmentations of FWER-controlling procedures, for control of the generalized family-wise error rate (Section 2.2) and the proportion of false positives among the rejected hypotheses (Section 2.3). Theorems 1 and 2 establish finite sample control and exact asymptotic control results for such augmentation procedures, under general data generating distributions, with arbitrary dependence structures among variables. Adjusted p-values for the gFWER and PFP-controlling augmentation procedures are shown to be simple functions of the adjusted p-values for the original FWER-controlling procedure. The companion articles give a detailed introduction to our general approach to multiple testing and provide single-step (Dudoit et al., 2003b; Pollard and van der Laan, 2003) and step-down (van der Laan et al., 2003) multiple testing procedures for controlling the FWER for general null hypotheses and test statistics. The basic framework and main definitions are recalled below for convenience.

2 Multiple testing procedures

2.1 Basic framework

Model. Let $X_1, \ldots, X_n$ be n independent and identically distributed (i.i.d.) random variables, $X \sim P \in \mathcal{M}$, where the *data generating distribution* P is known to be an element of a particular *statistical model* $\mathcal{M}$ (possibly non-parametric).

Null hypotheses. In order to cover a broad class of testing problems, we define m null hypotheses in terms of a collection of *submodels*, $\mathcal{M}_j \subseteq \mathcal{M}$, $j = 1, \ldots, m$, for the data generating distribution P. The m *null hypotheses* are defined as $H_{0j} \equiv \mathrm{I}(P \in \mathcal{M}_j)$ and the corresponding *alternative hypotheses* as $H_{1j} \equiv \mathrm{I}(P \notin \mathcal{M}_j)$.

For many testing problems of interest, each null hypothesis H_{0j} refers to a single parameter $\mu_j(P)$, $j = 1, \ldots, m$. The corresponding submodels are of the form $\mathcal{M}_j = \{P \in \mathcal{M} : \mu_j(P) = \mu_{0j}\}$, for two-sided tests, and $\mathcal{M}_j = \{P \in \mathcal{M} : \mu_j(P) \leq \mu_{0j}\}$ for one-sided tests, where μ_{0j} are user-supplied null-values.

Let $S_0 = S_0(P) \equiv \{j : H_{0j} \text{ is true}\} = \{j : P \in \mathcal{M}_j\}$ be the set of $m_0 = |S_0|$ true null hypotheses, where we note that S_0 depends on the true data generating distribution P. Let $S_0^c = \{1, \ldots, m\}/S_0 \equiv \{j : H_{0j} \text{ is false}\} = \{j : P \notin \mathcal{M}_j\}$ be the set of $m_1 = m - m_0$ false null hypotheses, i.e., true positives.

Multiple testing procedures. A *multiple testing procedure* (MTP) can be represented by a random subset S_n of $R_n = |S_n|$ rejected hypotheses, that *estimates* the set S_0^c of true positives,

$$S_n = S(T_n, Q_0, \alpha) \equiv \{j : H_{0j} \text{ is rejected}\} \subseteq \{1, \ldots, m\}. \tag{1}$$

As indicated by the long notation $S(T_n, Q_0, \alpha)$, the set S_n is a function of: (i) the data, $X_1, \ldots, X_n$, through an m-vector of test statistics $T_n = (T_n(j) : j = 1, \ldots, m)$, where each $T_n(j)$ corresponds to a null hypothesis H_{0j}; (ii) a test statistics null distribution, Q_0, for computing cut-offs for each $T_n(j)$ (and the resulting adjusted p-values); and (iii) the nominal level α of the MTP, i.e., the desired upper bound for a suitably defined Type I error rate. Multiple testing procedures such as those proposed in the companion articles (Dudoit et al., 2003b; van der Laan et al., 2003; Pollard and van der Laan, 2003) can be represented as

$$S_n = S(T_n, Q_0, \alpha) = \{j : T_n(j) > c_j\},$$

where $c_j = c_j(T_n, Q_0, \alpha)$, $j = 1, \ldots, m$, are possibly random *cut-offs*, or *critical values*, computed under the null distribution Q_0 for the test statistics.

Type I error rates. In any testing situation, two types of errors can be committed: a *false positive*, or *Type I error*, is committed by rejecting a true null hypothesis ($S_n \cap S_0$), and a *false negative*, or *Type II error*, is committed when the test procedure fails to reject a false null hypothesis ($S_n^c \cap S_0^c$). Denote the number of Type I errors by $V_n = |S_n \cap S_0|$. The multiple testing literature is concerned with constructing an optimal estimator of the set of true positives S_0^c (i.e., most powerful set, with minimal number of Type II

errors), while controlling a parameter $\theta_n = \theta(F_{V_n,R_n})$, of the joint distribution of the numbers of Type I errors V_n and rejected hypotheses R_n, at a user-supplied level α.

Here, we consider procedures that control the following two broad classes of Type I error rates: the generalized family-wise error rate and the proportion of false positives among the rejected hypotheses. The *generalized family-wise error rate* (gFWER), for a user-supplied integer $k = 0, \ldots, m_0 - 1$, is the probability of at least $(k+1)$ Type I errors. That is,

$$gFWER(k) \equiv Pr(V_n \geq k+1) = 1 - F_{V_n}(k), \tag{2}$$

where F_{V_n} is the discrete cumulative distribution function on $\{0, \ldots, m\}$ for the number of Type I errors, V_n. When $k = 0$, the gFWER is the usual *family-wise error rate* (FWER), or probability of at least one Type I error,

$$FWER \equiv Pr(V_n \geq 1) = 1 - F_{V_n}(0). \tag{3}$$

The *proportion of false positives among the rejected hypotheses* (PFP), for a user-supplied $q \in (0, 1)$, is defined as

$$PFP(q) \equiv Pr(V_n/R_n > q). \tag{4}$$

In the remainder of the article, we use the shorter phrase *proportion of false positives* to refer to the proportion of false positives *among the R_n rejected hypotheses* and not among the total number m of null hypotheses. Controlling the later proportion would amount to controlling the gFWER. Note that while the gFWER is a parameter of only the *marginal* distribution for the number of Type I errors V_n, the PFP is a parameter of the *joint* distribution of (V_n, R_n). Note also that the *false discovery rate* (FDR) is the *expected value* of the proportion of false positives,

$$FDR \equiv E[V_n/R_n] \tag{5}$$

(with the convention that $V_n/R_n \equiv 0$ if $R_n = 0$), while $PFP(q)$ is a *tail probability* for this proportion.

A multiple testing procedure $S_n = S(T_n, Q_0, \alpha)$ is said to provide *finite sample control* of a Type I error rate θ_n ($gFWER(k)$ or $PFP(q)$) at *level* $\alpha \in (0, 1)$, if $\theta_n \leq \alpha$, and *asymptotic control* of this error rate at level α, if $\limsup_n \theta_n \leq \alpha$. It is common practice to set $\alpha = 0.05$.

Adjusted p-values. Given a multiple testing procedure $S_n(\alpha) = S(T_n, Q_0, \alpha)$, the *adjusted p-value*, $\tilde{P}_n(j)$, for null hypothesis H_{0j}, is defined as

$$\tilde{P}_n(j) \equiv \inf\{\alpha : j \in S_n(\alpha)\}, \qquad j = 1, \ldots, m. \tag{6}$$

If $S_n(\alpha)$ is right-continuous at α, in the sense that $\lim_{\alpha' \downarrow \alpha} S_n(\alpha') = S_n(\alpha)$, then the adjusted p-values imply $S_n(\alpha)$ for each value of α through the relation

$$S_n(\alpha) = \{j : \tilde{P}_n(j) \leq \alpha\}. \tag{7}$$

Let $O_n(j)$ denote indices for the *ordered* adjusted p-values, so that $\tilde{P}_n(O_n(1)) \leq \ldots \leq \tilde{P}_n(O_n(m))$. Then, the set of rejected hypotheses $S_n(\alpha)$ consists of the indices for the $R_n(\alpha) = |S_n(\alpha)|$ hypotheses with the smallest adjusted p-values, that is, $S_n(\alpha) = \{O_n(j) : j = 1, \ldots, R_n(\alpha)\}$.

Augmentation procedures. This article proposes augmentations of FWER-controlling procedures that control the gFWER and PFP. That is, we presuppose a multiple testing procedure $S_n = S(T_n, Q_0, \alpha)$, that provides (asymptotic) control of the FWER at level α (Procedure 0, below). The reader is referred to the companion articles for a detailed discussion of the choice of test statistics T_n, null distribution Q_0, and specific proposals of single-step and step-down procedures that control the FWER (Dudoit et al., 2003b; van der Laan et al., 2003; Pollard and van der Laan, 2003). Accordingly, we take the test statistics T_n and their null distribution Q_0 as given, and denote the set of rejected hypotheses by $S_n(\alpha)$, to emphasize only the dependence of this set on the nominal level α of the MTP. Similarly, we denote the number of rejected hypotheses by $R_n(\alpha) \equiv |S_n(\alpha)|$ and the number of Type I errors by $V_n(\alpha) \equiv |S_n(\alpha) \cap S_0|$.

Procedure 0. FWER-controlling procedure.
Suppose one has available a multiple testing procedure $S_n(\alpha)$ that provides finite sample control of the FWER at level $\alpha_n \in (0, 1)$ and asymptotic control of the FWER at level $\alpha \in (0, 1)$. That is,

$$\alpha_n = Pr(V_n(\alpha) > 0), \qquad \forall n, \tag{8}$$

and

$$\limsup_{n \to \infty} Pr(V_n(\alpha) > 0) = \limsup_{n \to \infty} \alpha_n = \alpha^* \leq \alpha, \tag{9}$$

where $V_n(\alpha) = |S_n(\alpha) \cap S_0|$ denotes the number of Type I errors for procedure $S_n(\alpha)$.

Given a FWER-controlling procedure $S_n(\alpha)$ and a random subset $A_n(\alpha)$ (i.e., $A_n(\alpha)$ is a deterministic function of the data, $X_1, \ldots, X_n$), satisfying $Pr(A_n(\alpha) \subseteq S_n^c(\alpha)) = 1$, where $S_n^c(\alpha) \equiv \{1, \ldots, m\}/S_n(\alpha)$ is the set of non-rejected hypotheses, we define a new multiple testing procedure $S_n^+(\alpha)$ by

$$S_n^+(\alpha) \equiv S_n(\alpha) \cup A_n(\alpha). \tag{10}$$

That is, $S_n^+(\alpha)$ is an *augmentation* of the MTP $S_n(\alpha)$ by the set $A_n(\alpha)$, identifying $|A_n(\alpha)|$ additional rejections among the null hypotheses which were not rejected by $S_n(\alpha)$. In the sequel, the set $A_n(\alpha)$ will always denote a subset of $S_n^c(\alpha)$. We also define

$$R_n^+(\alpha) \equiv |S_n^+(\alpha)|,$$

as the number of rejected hypotheses, and

$$V_n^+(\alpha) \equiv |S_n^+(\alpha) \cap S_0|,$$

as the number of Type I errors for the *augmentation multiple testing procedure* $S_n^+(\alpha)$.

Sections 2.2 and 2.3, below, provide explicit proposals of augmentation sets $A_n(\alpha)$ that define procedures $S_n^+(\alpha)$ for controlling the gFWER and the PFP, respectively. Theorems 1 and 2 establish finite sample control and exact asymptotic control results for such augmentation procedures.

2.2 Multiple testing procedures for control of the gFWER

2.2.1 Augmentation procedures

For control of the gFWER, Theorem 1 states that, for any non-negative integer k and set $A_n(\alpha)$ with $Pr(|A_n(\alpha)| = k) = 1$, the multiple testing procedure $S_n^+(\alpha) = S_n(\alpha) \cup A_n(\alpha)$ provides finite sample control of $gFWER(k)$ at level α_n, that is, $Pr(V_n^+(\alpha) > k) \leq \alpha_n$, $\forall n$. In addition, if $S_n(\alpha)$ controls the FWER asymptotically exactly at level α^* and the true alternatives are asymptotically always rejected by $S_n(\alpha)$ (condition (14), below), then the augmentation procedure $S_n^+(\alpha)$ also provides asymptotically exact control of $gFWER(k)$ at level α^*, that is, $\liminf_n Pr(V_n^+(\alpha) = k) = 1 - \alpha^*$.

Though Theorem 1 applies to any random set $A_n(\alpha)$ satisfying the specified size constraints with probability one, because of power considerations, we

propose the following construction for $A_n(\alpha)$, based on the ordered adjusted p-values for the FWER-controlling procedure.

Procedure 1. Augmentation procedure for control of the gFWER. Consider a multiple testing procedure $S_n(\alpha)$, such as Procedure 0, that provides finite sample control of the FWER at level α_n and asymptotic control of the FWER at level α.

1. First, order the m hypotheses according to their FWER adjusted p-values, $\tilde{P}_n(j)$, from smallest to largest, that is, define indices $O_n(j)$, so that $\tilde{P}_n(O_n(1)) \leq \ldots \leq \tilde{P}_n(O_n(m))$. The FWER-controlling procedure rejects null hypotheses

$$S_n(\alpha) = \{j : \tilde{P}_n(j) \leq \alpha\} = \{O_n(j) : j = 1, \ldots, R_n(\alpha)\}. \tag{11}$$

2. For a given integer $k = 0, \ldots, m - R_n(\alpha)$, define an augmentation gFWER-controlling procedure $S_n^+(\alpha)$ by

$$S_n^+(\alpha) \equiv S_n(\alpha) \cup A_n(k, \alpha), \tag{12}$$

where $A_n(k, \alpha)$ is an augmentation set of size k defined by

$$A_n(k, \alpha) \equiv \{O_n(j) : j = R_n(\alpha) + 1, \ldots, R_n(\alpha) + k\}. \tag{13}$$

That is, the set $A_n(k, \alpha)$ corresponds to the k most significant hypotheses that were *not* rejected by the FWER-controlling procedure $S_n(\alpha)$.

2.2.2 Finite sample and asymptotic control of gFWER

Theorem 1 [Control of gFWER] *Consider a multiple testing procedure $S_n(\alpha)$, such as Procedure 0, that provides finite sample control of the FWER at level α_n and asymptotic control of the FWER at level α. That is, $Pr(V_n(\alpha) > 0) = \alpha_n$, $\forall n$, and $\limsup_n Pr(V_n(\alpha) > 0) = \limsup_n \alpha_n = \alpha^* \leq \alpha$.*

Finite sample control. *For a non-negative integer k, suppose $A_n(\alpha)$ is a random subset such that the events $A_n(\alpha) \subseteq S_n^c(\alpha)$ and $|A_n(\alpha)| = \min(k, m - |S_n(\alpha)|)$ have joint probability one. Then, the augmentation procedure $S_n^+(\alpha) = S_n(\alpha) \cup A_n(\alpha)$ provides finite sample control of $gFWER(k)$ at level α_n, that*

is,

$$Pr(V_n^+(\alpha) > k) \leq Pr(V_n(\alpha) > 0) = \alpha_n, \qquad \forall n.$$

Asymptotic exact control. *Consider a random subset $A_n(\alpha)$ such that the events $A_n(\alpha) \subseteq S_n^c(\alpha)$ and $|A_n(\alpha)| = k$ have joint probability one. Asymptotic control at level α follows immediately from finite sample control above. Suppose now that $S_n(\alpha)$ also satisfies*

$$\lim_{n\to\infty} Pr(S_0^c \subseteq S_n(\alpha)) = 1, \tag{14}$$

that is, the true alternative hypotheses are asymptotically always rejected by $S_n(\alpha)$. Then, the augmentation procedure $S_n^+(\alpha) = S_n(\alpha) \cup A_n(\alpha)$ is such that

$$\liminf_{n\to\infty} Pr(V_n^+(\alpha) = k) = 1 - \alpha^*. \tag{15}$$

In particular, $S_n^+(\alpha)$ provides exact asymptotic control of $gFWER(k)$ at level $\alpha^* \leq \alpha$, that is,

$$\limsup_{n\to\infty} Pr(V_n^+(\alpha) > k) = \alpha^*.$$

Note that the augmentation set $A_n(k, \alpha)$ in Procedure 1 trivially satisfies the conditions of Theorem 1.

Proof of Theorem 1.

Finite sample control. Since $Pr(|A_n(\alpha)| \leq k) = 1$, we have $Pr(V_n^+(\alpha) \leq V_n(\alpha) + k) = 1$. Thus, $Pr(V_n^+(\alpha) > k) \leq Pr(V_n(\alpha) + k > k) = Pr(V_n(\alpha) > 0)$, which equals α_n by definition.

Asymptotic exact control. Define $B_n \equiv \mathrm{I}(S_0^c \subseteq S_n(\alpha))$. By assumption, $\lim_n Pr(B_n = 1) = 1$. Thus,

$$Pr(V_n^+(\alpha) = k) = Pr(V_n^+(\alpha) = k \mid B_n = 1)Pr(B_n = 1) + o(1).$$

Note that, if $B_n = 1$, then $A_n(\alpha) \subseteq S_n^c(\alpha) \subseteq S_0$, so that $V_n^+(\alpha) = V_n(\alpha) + k$, where we use that $|A_n(\alpha)| = k$ with probability one. Thus,

$$\begin{aligned} Pr(V_n^+(\alpha) = k) &= Pr(V_n(\alpha) + k = k \mid B_n = 1)Pr(B_n = 1) + o(1) \\ &= Pr(V_n(\alpha) = 0 \mid B_n = 1)Pr(B_n = 1) + o(1) \\ &= Pr(V_n(\alpha) = 0) - Pr(V_n(\alpha) = 0, B_n = 0) + o(1) \\ &= Pr(V_n(\alpha) = 0) + o(1), \end{aligned}$$

where the last equality follows by noting that $Pr(V_n(\alpha) = 0, B_n = 0) \leq Pr(B_n = 0) \to 0$, as $n \to \infty$. By assumption, we have $\liminf_n Pr(V_n(\alpha) = 0) = 1 - \alpha^*$, thus, as required $\liminf_n Pr(V_n^+(\alpha) = k) = 1 - \alpha^*$.

□

2.2.3 Adjusted p-values

Consider multiple testing Procedure 1, for controlling $gFWER(k)$ at level α, where $S_n^+(\alpha) = S_n(\alpha) \cup A_n(k, \alpha)$ and the augmentation set $A_n(k, \alpha)$ is defined in equation (13). By definition, the adjusted p-values $\tilde{P}_n^+(j)$ for augmentation procedure $S_n^+(\alpha)$ are given by

$$\tilde{P}_n^+(j) = \inf\{\alpha : j \in S_n^+(\alpha)\}.$$

As one might expect, these adjusted p-values are trivial functions of the adjusted p-values $\tilde{P}_n(j)$ for the FWER-controlling procedure $S_n(\alpha)$. Recall the ordering $H_{0,O_n(1)}, \ldots, H_{0,O_n(m)}$ of the null hypotheses, with corresponding FWER adjusted p-values $\tilde{P}_n(O_n(1)) \leq \ldots \leq \tilde{P}_n(O_n(m))$. Then, the gFWER adjusted p-values are given by

$$\tilde{P}_n^+(O_n(j)) = \begin{cases} 0, & \text{if } j = 1, \ldots, k, \\ \tilde{P}_n(O_n(j-k)), & \text{if } j = k+1, \ldots, m. \end{cases} \tag{16}$$

2.3 Multiple testing procedures for control of the PFP

2.3.1 Augmentation procedures

For control of the PFP, given $q \in (0, 1)$, define

$$k_n(q, \alpha) \equiv \max\left\{ j \in \{0, \ldots, m - R_n(\alpha)\} : \frac{j}{j + R_n(\alpha)} \leq q \right\}$$

and

$$q^* \equiv \frac{k_n(q, \alpha)}{k_n(q, \alpha) + R_n(\alpha)} \leq q, \tag{17}$$

where $R_n(\alpha) = |S_n(\alpha)|$ denotes the number of rejected hypotheses for FWER-controlling procedure $S_n(\alpha)$. Then, Theorem 2 states that, for a set $A_n(\alpha)$ with $Pr(|A_n(\alpha)| = k_n(q, \alpha)) = 1$, the multiple testing procedure $S_n^+(\alpha) =$

$S_n(\alpha) \cup A_n(\alpha)$ provides finite sample control of $PFP(q^*)$ at level α_n, that is, $Pr\left(V_n^+(\alpha)/R_n^+(\alpha) > q^*\right) \leq \alpha_n, \forall n$. In addition, if $S_n(\alpha)$ controls the FWER asymptotically exactly at level α^* and the true alternatives are asymptotically always rejected by $S_n(\alpha)$ (condition (14), in Theorem 1), then the augmentation procedure $S_n^+(\alpha)$ also provides asymptotically exact control of $PFP(q^*)$ at level α^*, that is, $\liminf_n Pr(V_n^+(\alpha)/R_n^+(\alpha) = q^*) = 1 - \alpha^*$.

As does Theorem 1 for gFWER control, Theorem 2 applies to any random set $A_n(\alpha)$ satisfying the specified size constraints with probability one. However, for power considerations, we propose the following specific construction for the augmentation set $A_n(\alpha)$, based on the ordered adjusted p-values for the FWER-controlling procedure.

Procedure 2. Augmentation procedure for control of the PFP. Consider a multiple testing procedure $S_n(\alpha)$, such as Procedure 0, that provides finite sample control of the FWER at level α_n and asymptotic control of the FWER at level α.

1. First, order the m hypotheses according to their FWER adjusted p-values, $\tilde{P}_n(j)$, from smallest to largest, that is, define indices $O_n(j)$, so that $\tilde{P}_n(O_n(1)) \leq \ldots \leq \tilde{P}_n(O_n(m))$. The FWER-controlling procedure rejects null hypotheses

$$S_n(\alpha) = \{j : \tilde{P}_n(j) \leq \alpha\} = \{O_n(j) : j = 1, \ldots, R_n(\alpha)\}. \tag{18}$$

2. For a given $q \in (0, 1)$, define an augmentation PFP-controlling procedure $S_n^+(\alpha)$ by

$$S_n^+(\alpha) \equiv S_n(\alpha) \cup A_n(k_n(q, \alpha), \alpha), \tag{19}$$

where $A_n(k_n(q, \alpha), \alpha)$ is an augmentation set of size $k_n(q, \alpha)$ defined by

$$k_n(q, \alpha) \equiv \max \left\{ j \in \{0, \ldots, m - R_n(\alpha)\} : \frac{j}{j + R_n(\alpha)} \leq q \right\} \tag{20}$$

and

$$A_n(k_n(q, \alpha), \alpha) \equiv \{O_n(j) : j = R_n(\alpha)+1, \ldots, R_n(\alpha)+k_n(q, \alpha)\}. \tag{21}$$

That is, the set $A_n(k_n(q, \alpha), \alpha)$ corresponds to the $k_n(q, \alpha)$ most significant hypotheses that were *not* rejected by the FWER-controlling procedure $S_n(\alpha)$.

2.3.2 Finite sample and asymptotic control of PFP

Theorem 2 [Control of PFP] *Consider a multiple testing procedure $S_n(\alpha)$, such as Procedure 0, that provides finite sample control of the FWER at level α_n and asymptotic control of the FWER at level α. That is, $Pr(V_n(\alpha) > 0) = \alpha_n$, $\forall n$, and $\limsup_n Pr(V_n(\alpha) > 0) = \limsup_n \alpha_n = \alpha^* \leq \alpha$. For $q \in (0, 1)$, suppose $A_n(\alpha)$ is a random subset such that the events $A_n(\alpha) \subseteq S_n^c(\alpha)$ and $|A_n(\alpha)| = k_n(q, \alpha)$ have joint probability one.*

Finite sample control. *The augmentation procedure* $S_n^+(\alpha) = S_n(\alpha) \cup A_n(\alpha)$ *provides finite sample control of* $PFP(q^*)$ *at level* α_n*, that is,*

$$Pr\left(\frac{V_n^+(\alpha)}{R_n^+(\alpha)} > q^*\right) \leq Pr(V_n(\alpha) > 0) = \alpha_n, \qquad \forall n.$$

Asymptotic exact control. *Asymptotic control at level* α *follows immediately from finite sample control above. Suppose now that* $S_n(\alpha)$ *also satisfies condition (14), that is,* $\lim_n Pr(S_0^c \subseteq S_n(\alpha)) = 1$*. Then, the augmentation procedure* $S_n^+(\alpha) = S_n(\alpha) \cup A_n(\alpha)$ *is such that*

$$\liminf_{n\to\infty} Pr\left(\frac{V_n^+(\alpha)}{R_n^+(\alpha)} = q^*\right) = 1 - \alpha^*. \tag{22}$$

In particular, $S_n^+(\alpha)$ *provides exact asymptotic control of* $PFP(q^*)$ *at level* $\alpha^* \leq \alpha$*, that is,*

$$\limsup_{n\to\infty} Pr\left(\frac{V_n^+(\alpha)}{R_n^+(\alpha)} > q^*\right) = \alpha^*.$$

Note that the augmentation set $A_n(k_n(q,\alpha),\alpha)$ in Procedure 2 trivially satisfies the conditions of Theorem 2. Unlike the PFP-controlling procedures in Genovese and Wasserman (2001), which assume independence of the test statistics, our proposal controls the PFP for test statistics with arbitrary joint distributions.

Proof of Theorem 2.

Finite sample control. Since $Pr(|A_n(\alpha)| = k_n(q,\alpha)) = 1$, we have $V_n^+(\alpha) \leq V_n(\alpha) + k_n(q,\alpha)$ and $R_n^+(\alpha) = R_n(\alpha) + k_n(q,\alpha)$. Thus,

$$\frac{V_n^+(\alpha)}{R_n^+(\alpha)} \leq \frac{V_n(\alpha) + k_n(q,\alpha)}{R_n(\alpha) + k_n(q,\alpha)}.$$

Define the indicator $C_n \equiv \mathrm{I}(V_n(\alpha) = 0)$. Given $C_n = 1$ and by definition of q^*, we have that

$$\frac{V_n(\alpha) + k_n(q,\alpha)}{R_n(\alpha) + k_n(q,\alpha)} = \frac{k_n(q,\alpha)}{R_n(\alpha) + k_n(q,\alpha)} = q^*.$$

Thus,

$$\begin{aligned} Pr\left(\frac{V_n^+(\alpha)}{R_n^+(\alpha)} > q^*\right) &\leq Pr\left(\frac{V_n(\alpha)+k_n(q,\alpha)}{R_n(\alpha)+k_n(q,\alpha)} > q^*\right) \\ &= Pr(\cdot > q^* | C_n = 1) Pr(C_n = 1) \\ &\quad + Pr(\cdot > q^* \mid C_n = 0) Pr(C_n = 0) \\ &= 0 * Pr(C_n = 1) + Pr(\cdot > q^* \mid C_n = 0)\alpha_n \\ &\leq \alpha_n, \end{aligned}$$

where we use the short-hand notation $\cdot$ for $(V_n(\alpha) + k_n(q,\alpha))/(R_n(\alpha) + k_n(q,\alpha))$.

Exact asymptotic control. As in the proof of Theorem 1, let $B_n \equiv \mathrm{I}(S_0^c \subseteq S_n(\alpha))$. Given $B_n = 1$, we have $V_n^+(\alpha) = V_n(\alpha) + k_n(q,\alpha)$ and hence

$$\frac{V_n^+(\alpha)}{R_n^+(\alpha)} = \frac{V_n(\alpha)+k_n(q,\alpha)}{R_n(\alpha)+k_n(q,\alpha)}. \tag{23}$$

Recall the definition $C_n \equiv \mathrm{I}(V_n(\alpha) = 0)$ and also define $D_n \equiv \mathrm{I}(B_n = C_n = 1)$. Since, by assumption, $\lim_n Pr(B_n = 1) = 1$ and $\liminf_n Pr(C_n = 1) = 1 - \alpha^*$, we have $\liminf_n Pr(D_n = 1) = 1 - \alpha^*$. Given $D_n = 1$ (i.e., $V_n(\alpha) = 0$ and $A_n(\alpha) \subseteq S_n^c(\alpha) \subseteq S_0$) and by definition of q^*, we have that

$$\frac{V_n(\alpha)+k_n(q,\alpha)}{R_n(\alpha)+k_n(q,\alpha)} = \frac{k_n(q,\alpha)}{R_n(\alpha)+k_n(q,\alpha)} = q^*.$$

Thus,

$$\begin{aligned} Pr\left(\frac{V_n^+(\alpha)}{R_n^+(\alpha)} = q^*\right) &= Pr\left(\frac{V_n^+(\alpha)}{R_n^+(\alpha)} = q^* \mid D_n = 1\right) Pr(D_n = 1) \\ &\quad + Pr\left(\frac{V_n^+(\alpha)}{R_n^+(\alpha)} = q^* \mid D_n = 0\right) Pr(D_n = 0) \\ &= 1 * Pr(D_n = 1) \\ &\quad + Pr\left(\frac{V_n^+(\alpha)}{R_n^+(\alpha)} = q^* \mid D_n = 0\right) Pr(D_n = 0) \end{aligned}$$

and hence

$$\begin{aligned} &\liminf_{n\to\infty} Pr\left(\tfrac{V_n^+(\alpha)}{R_n^+(\alpha)} = q^*\right) \\ &\quad = (1-\alpha^*) + \liminf_{n\to\infty} Pr\left(\tfrac{V_n^+(\alpha)}{R_n^+(\alpha)} = q^* \mid D_n = 0\right) Pr(D_n = 0). \end{aligned}$$

In the remainder of the proof, we show that the second term equals zero. The event $\{D_n = 0\}$ is a union of three events, $\{B_n = 1, C_n = 0\}$, $\{B_n = 0, C_n = 1\}$, and $\{B_n = 0, C_n = 0\}$. Because $\lim_n Pr(B_n = 0) = 0$, it follows that the last two events happen with probability tending to zero. Consequently, we have

$$\begin{aligned} &Pr\left(\tfrac{V_n^+(\alpha)}{R_n^+(\alpha)} = q^* \mid D_n = 0\right) Pr(D_n = 0) \\ &\quad = Pr\left(\tfrac{V_n^+(\alpha)}{R_n^+(\alpha)} = q^* \mid B_n = 1, C_n = 0\right) Pr(B_n = 1, C_n = 0) + o(1). \end{aligned}$$

Finally, by (23), we have

$$\begin{aligned} Pr\left(\tfrac{V_n^+(\alpha)}{R_n^+(\alpha)} = q^* \mid B_n = 1, C_n = 0\right) &= Pr\left(\tfrac{V_n(\alpha)+k_n(q,\alpha)}{R_n(\alpha)+k_n(q,\alpha)} = q^* \mid B_n = 1, C_n = 0\right) \\ &= Pr(V_n(\alpha) = 0 \mid B_n = 1, C_n = 0) \\ &= 0. \end{aligned}$$

This completes the proof that $\liminf_n Pr(V_n^+(\alpha)/R_n^+(\alpha) = q^*) = 1 - \alpha^*$.

□

2.3.3 Adjusted p-values

Consider multiple testing Procedure 2, for controlling $PFP(q^*)$ at level α, where $S_n^+(\alpha) = S_n(\alpha) \cup A_n(k_n(q,\alpha),\alpha)$ corresponds to the augmentation set $A_n(k_n(q,\alpha),\alpha)$ defined in equation (21). Let

$$U_n(\alpha) \equiv \frac{1}{m} \sum_{j=1}^{m} \mathrm{I}(\tilde{P}_n(j) \leq \alpha), \tag{24}$$

where $\tilde{P}_n(j)$ are the adjusted p-values for the FWER-controlling procedure $S_n(\alpha)$. For simplicity, suppose $S_n(\alpha)$ is right-continuous at α, for all $\alpha \in (0,1)$, so that we have the representation $S_n(\alpha) = \{j : \tilde{P}_n(j) \leq \alpha\}$ and thereby $R_n(\alpha) = \sum_{j=1}^m \mathrm{I}(\tilde{P}_n(j) \leq \alpha) = U_n(\alpha) * m$. Let $O_n(j)$ denote indices for the ordered adjusted p-values, $\tilde{P}_n(O_n(j))$, for FWER-controlling procedure $S_n(\alpha)$, so that $\tilde{P}_n(O_n(1)) \leq \ldots \leq \tilde{P}_n(O_n(m))$. Then, the adjusted p-values, $\tilde{P}_n^+(O_n(j))$, corresponding to null hypotheses $H_{0,O_n(j)}$, for the PFP-

controlling procedure $S_n^+(\alpha)$, are given by

$$\begin{aligned}
\tilde{P}_n^+(O_n(j)) &= \inf\{\alpha : O_n(j) \in S_n^+(\alpha)\} \\
&= \inf\left\{\alpha : \frac{j - R_n(\alpha)}{j} \leq q\right\} \\
&= \inf\left\{\alpha : \frac{j - U_n(\alpha) * m}{j} \leq q\right\} \\
&= \inf\left\{\alpha : U_n(\alpha) \geq (1-q)\frac{j}{m}\right\} \\
&= U_n^{-1}\left((1-q)\frac{j}{m}\right), \qquad j = 1, \ldots, m. \qquad (25)
\end{aligned}$$

Here, we define the inverse F^{-1} of a cumulative distribution function F by $F^{-1}(y) \equiv \inf\{x : F(x) \geq y\}$, and note that $F(x) \geq y \Leftrightarrow x \geq F^{-1}(y)$. Thus, our proposed augmentation multiple testing procedure $S_n^+(\alpha)$, for controlling $PFP(q^*)$ at level α, can be stated in terms of the adjusted p-values for the FWER-controlling procedure $S_n(\alpha)$, as

$$\begin{aligned}
S_n^+(\alpha) &= \left\{O_n(j) : U_n^{-1}\left((1-q)\frac{j}{m}\right) \leq \alpha\right\} \\
&= \left\{O_n(j) : j \leq \frac{\sum_{l=1}^m \mathrm{I}(\tilde{P}_n(l) \leq \alpha)}{1-q}\right\}. \qquad (26)
\end{aligned}$$

2.4 Available FWER-controlling multiple testing procedures

Consider an m-vector of test statistics $T_n = (T_n(j) : j = 1, \ldots, m)$, where each $T_n(j)$ corresponds to a null hypothesis H_{0j}, and let $Q_n = Q_n(P)$ denote the joint distribution of these test statistics. Multiple testing procedures available in the literature for (asymptotic) control of the FWER at level α are of the form

$$S_n = S(T_n, Q_0, \alpha) = \{j : T_n(j) > c_j\},$$

where $c_j = c_j(T_n, Q_0, \alpha)$, $j = 1, \ldots, m$, are possibly random cut-offs, computed under a null distribution Q_0 for the test statistics. As detailed in the companion articles (Dudoit et al., 2003b; van der Laan et al., 2003; Pollard and van der Laan, 2003), different proposals of MTPs correspond with

different cut-off rules, e.g., single-step vs. step-down, common-quantile vs. common-cut-off procedures.

These three articles provide specific proposals for single-step and step-down common-quantile and common-cut-off procedures that asymptotically control the FWER, for general null hypotheses and test statistics. A key feature of these procedures is the test statistics null distribution Q_0 (rather than data generating null distribution) used to derive the cut-offs c_j (and the resulting adjusted p-values). Dudoit et al. (2003b) and van der Laan et al. (2003) identify an asymptotic *null domination condition*, whereby the number of Type I errors, $V_n(\alpha) = |S_n(\alpha) \cap S_0|$, is asymptotically stochastically larger under the test statistics null distribution Q_0 than under the true distribution $Q_n(P)$. As a result, asymptotic control of the Type I error rate under the assumed null distribution does indeed imply asymptotic control of this error rate under the true, unknown data generating distribution. When the true distribution of the S_0-specific test statistics asymptotically equals their null distribution under Q_0, one has asymptotic exact control of the Type I error rate. Dudoit et al. (2003b) and van der Laan et al. (2003) propose as an explicit null distribution the asymptotic distribution of the vector of null-value shifted and scaled test statistics and provide a general bootstrap algorithm to conveniently obtain consistent estimators of this distribution. Single-step and step-down procedures based on consistent estimators of the joint null distribution are shown to also provide asymptotic control of the Type I error rate, for general data generating distributions, without the need for conditions such as subset pivotality.

Alternatively, one can use a conservative explicit bound on the tail probability of the number of rejections, under a proper null distribution such as that proposed in Dudoit et al. (2003b) and van der Laan et al. (2003). This bound is typically based on the tail probabilities of the *marginal* distributions of the test statistics $(T_n(j) : j \in S_0)$ for the true null hypotheses and relies on standard probability inequalities, such as Boole's inequality: $Pr(\cup_{l=1}^{m} B_l) \leq \sum_{l=1}^{m} Pr(B_l)$, for events B_l of the form $B_l = \{T_n(l) > c_l\}$. While such an approach circumvents the need for resampling techniques to estimate the joint distribution of the vector of test statistics, it can result in very conservative multiple testing procedures. Examples of such methods are single-step and step-down procedures based on known marginal distributions for the test statistics (e.g., $N(0, 1)$) or on finite sample conservative marginal tail probabilities as one obtains from using Bernstein's and Boole's inequalities (Hochberg, 1988; Holm, 1979; Hommel, 1988; Rom, 1990; Simes,

1986).

2.5 Multiple testing procedures for control of the FDR

Consider a multiple testing procedure $S_n(\alpha)$, such as Procedure 0, asymptotically controlling the FWER at level α. To obtain our first (more refined) result regarding control of the false discovery rate, we make the slightly stronger assumptions that the limit of FWER exists, that is, $\lim_n Pr(V_n(\alpha) > 0) = \alpha^* \leq \alpha$, and that the true alternative hypotheses are asymptotically always rejected by $S_n(\alpha)$ (condition (14), in Theorem 1, above). The first assumption implies that the statements of Theorem 2 hold with $\limsup_n$ and $\liminf_n$ replaced by $\lim_n$. Next, consider a corresponding PFP augmentation procedure $S_n^+(\alpha)$, such as Procedure 2, with given q and α. Then, by Theorem 2, procedure $S_n^+(\alpha)$ satisfies

$$\lim_{n\to\infty} Pr\left(\frac{V_n^+(\alpha)}{R_n^+(\alpha)} = q^*\right) = 1-\alpha^* \text{ and } \lim_{n\to\infty} Pr\left(\frac{V_n^+(\alpha)}{R_n^+(\alpha)} > q^*\right) = \alpha^*,$$

where q^* is defined in equation (17) and is such that $q^* \leq q$. Consequently, one can bound the false discovery rate (FDR) of augmentation procedure $S_n^+(\alpha)$ as follows

$$\begin{aligned}
\lim_{n\to\infty} E\left[\frac{V_n^+(\alpha)}{R_n^+(\alpha)}\right] &= \lim_{n\to\infty} E\left[\frac{V_n^+(\alpha)}{R_n^+(\alpha)}\middle|\, \frac{V_n^+(\alpha)}{R_n^+(\alpha)} \leq q^*\right] Pr\left(\frac{V_n^+(\alpha)}{R_n^+(\alpha)} \leq q^*\right) \\
&\quad + \lim_{n\to\infty} E\left[\frac{V_n^+(\alpha)}{R_n^+(\alpha)}\middle|\, \frac{V_n^+(\alpha)}{R_n^+(\alpha)} > q^*\right] Pr\left(\frac{V_n^+(\alpha)}{R_n^+(\alpha)} > q^*\right) \\
&\leq q^*(1-\alpha^*) + \alpha^* \\
&\leq q(1-\alpha^*) + \alpha^*, \qquad (27)
\end{aligned}$$

where the second to last inequality follows from replacing the conditional expectation $E\left[V_n^+(\alpha)/R_n^+(\alpha) | V_n^+(\alpha)/R_n^+(\alpha) > q^*\right]$ by the conservative upper bound of one and the last inequality follows from $q^* \leq q$. Note that the above bound is non-parametric (i.e., it does not rely on independence assumptions). It is conservative, since it corresponds to a distribution for which, in the limit, $Pr(V_n^+(\alpha)/R_n^+(\alpha) = 1) = \alpha^*$. In the situation that $\alpha^* = \alpha$ (i.e., for exact asymptotic control, as can be obtained with the step-down FWER-controlling procedures in van der Laan et al. (2003)), then setting, for example, $\alpha = q = 1 - \sqrt{1-q_1}$, results in asymptotic control of the FDR at level q_1, i.e., in $\lim_n E\left[V_n^+(\alpha)/R_n^+(\alpha)\right] \leq q_1$.

In general, for any procedure $S_n(\alpha)$ asymptotically controlling the FWER at level α, that is, such that $\limsup_n Pr(V_n(\alpha) > 0) = \alpha^* \leq \alpha$, Theorem 2 implies the following more conservative bound for the FDR of the corresponding augmentation PFP procedure $S_n^+(\alpha)$, for any user-supplied q and α,

$$\limsup_{n\to\infty} E\left[\frac{V_n^+(\alpha)}{R_n^+(\alpha)}\right] \leq q + \alpha. \tag{28}$$

Hence, setting, for example, $\alpha = q = q_1/2$, results in asymptotic control of the FDR at level q_1. The latter bound is conservative in two ways: from bounding $\limsup_n E\left[V_n^+(\alpha)/R_n^+(\alpha) | V_n^+(\alpha)/R_n^+(\alpha) > q\right]$ by one, as above, and also from replacing $\limsup_n Pr(V_n^+(\alpha)/R_n^+(\alpha) \leq q)$ by the conservative upper bound of one. Note that the finite sample result in Theorem 2 implies that the FDR of procedure $S_n^+(\alpha)$ is bounded by $q + \alpha_n$. The above arguments provide us with the following two procedures for asymptotically controlling the FDR at a user-supplied level q_1.

Theorem 3 [Control of FDR at level q_1] *Consider a multiple testing procedure $S_n(\alpha)$ asymptotically controlling the FWER at level α, that is, such that $\limsup_n Pr(V_n(\alpha) > 0) = \alpha^* \leq \alpha$. Then, the corresponding PFP augmentation procedure $S_n^+(\alpha)$ (e.g., Procedure 2), with $\alpha = q = q_1/2$, asymptotically controls the FDR at level q_1, that is,*

$$\limsup_{n\to\infty} E\left[\frac{V_n^+(\alpha)}{R_n^+(\alpha)}\right] \leq q_1. \tag{29}$$

If, in addition, (i) $\alpha^* = \alpha$, (ii) $\lim_n Pr(V_n(\alpha) > 0) = \alpha^*$, and (iii) $\lim_n Pr(S_0^c \subseteq S_n(\alpha)) = 1$ (condition (14)), then augmentation procedure $S_n^+(\alpha)$, with $\alpha = q = 1 - \sqrt{1 - q_1}$, asymptotically controls the FDR at level q_1.

It would be interesting to compare these new proposals to the conservative procedure of Benjamini and Yekutieli (2001), which is, to our knowledge, the only fully non-parametric FDR-controlling procedure in the literature.

3 Conclusions

The present article aimed to fill a gap in the current multiple testing literature by providing generally applicable multiple testing procedures (sharply)

controlling the number and proportion of false positives. Specifically, Theorems 1 and 2 teach us that one can map a multiple testing procedure $S_n(\alpha)$, controlling the FWER at level α_n (for a fixed sample size n), directly and computationally trivially, into multiple testing procedures $S_n^+(\alpha)$ controlling $gFWER(k)$ and $PFP(q)$ at level α_n, for any user-supplied $k = 0, 1, \ldots$ and $q \in (0, 1)$. Moreover, if as n converges to infinity, (a) the multiple testing procedure $S_n(\alpha)$ provides exact asymptotic control of the FWER at level α and (b) $S_n(\alpha)$ contains all true positives with probability tending to one, then the corresponding augmented multiple testing procedures $S_n^+(\alpha)$ also asymptotically *exactly* control $gFWER(k)$ and $PFP(q)$ at level α. Conditions (a) and (b) can be achieved, in general, with the resampling based step-down procedures presented in van der Laan et al. (2003). Finally, for any user-supplied k or q, the adjusted p-values for the augmented multiple testing procedures $S_n^+(\alpha)$ are trivial functions of the ordered FWER-adjusted p-values for $S_n(\alpha)$ (equations (16) and (25)). The important practical implication of the latter is that a multiple testing procedure $S_n(\alpha)$ (asymptotically) controlling FWER at level α, and its corresponding adjusted p-values, provide us without additional work with multiple testing procedures controlling $gFWER(k)$ and $PFP(q)$ at level α for any $k = 0, 1, \ldots$, $q \in (0, 1)$, and $\alpha \in (0, 1)$.

The multiple testing procedures proposed in this and accompanying articles (Dudoit et al., 2003b; van der Laan et al., 2003; Pollard and van der Laan, 2003) will be implemented in the near future in the open source R package `multtest`, released as part of the Bioconductor Project (`www.bioconductor.org`). A large scale simulation study investigating the practical performance of these procedures is presented in (Pollard et al., 2004).

References

F. Abramovich, Y. Benjamini, D. Donoho, and I. Johnstone. Adapting to unknown sparsity by controlling the false discovery rate. Technical report, Department of Statistics, Stanford University, 2000.

Y. Benjamini and H. Braun. John W. Tukey's contributions to multiple comparisons. *Annals of Statistics*, 30(6):1576–1594, 2002.

Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a prac-

tical and powerful approach to multiple testing. *J. R. Statist. Soc. B*, 57: 289–300, 1995.

Y. Benjamini and D. Yekutieli. The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, 29(4):1165–1188, 2001.

S. Dudoit, J. P. Shaffer, and J. C. Boldrick. Multiple hypothesis testing in microarray experiments. *Statistical Science*, 18(1):71–103, 2003a.

S. Dudoit, M. J. van der Laan, and K. S. Pollard. Multiple testing. Part I. Single-step procedures for control of general Type I error rates. Technical Report 138, Division of Biostatistics, UC Berkeley, 2003b. URL `www.bepress.com/ucbbiostat/paper138/`.

B. Efron, J. D. Storey, and R. Tibshirani. Microarrays, empirical Bayes methods, and false discovery rates. Technical Report 2001-218, Department of Statistics, Stanford University, 2001a.

B. Efron, R. Tibshirani, J. D. Storey, and V. Tusher. Empirical Bayes analysis of a microarray experiment. *Journal of the American Statistical Association*, 96:1151–1160, 2001b.

C. Genovese and L. Wasserman. Operating characteristics and extensions of the FDR procedure. Technical Report 737, Department of Statistics, Carnegie Mellon, 2001.

C. Genovese and L. Wasserman. False discovery rates. Technical Report 762, Department of Statistics, Carnegie Mellon, 2002.

Y. Hochberg. A sharper Bonferroni procedure for multiple tests of significance. *Biometrika*, 75:800–802, 1988.

Y. Hochberg and A. C. Tamhane. *Multiple Comparison Procedures*. Probability and Mathematical Statistics. Wiley–Interscience, 1987.

S. Holm. A simple sequentially rejective multiple test procedure. *Scand. J. Statist.*, 6:65–70, 1979.

G. Hommel. A stagewise rejective multiple test procedure based on a modified Bonferroni test. *Biometrika*, 75:383–386, 1988.

K. S. Pollard, M. D. Birkner, S. Dudoit, and M. J. van der Laan. Multiple testing. Part IV. Assessment of multiple testing procedures: Simulation studies and applications to genomic data analysis. Technical report, Division of Biostatistics, UC Berkeley, 2004. (In preparation).

K. S. Pollard and M. J. van der Laan. Resampling-based multiple testing: Asymptotic control of Type I error and applications to gene expression data. Technical Report 121, Department of Biostatistics, University of California, Berkeley, 2003. URL `www.bepress.com/ucbbiostat/paper121/`. (To appear in the Journal of Statistical Planning and Inference).

A. Reiner, D. Yekutieli, and Y. Benjamini. Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics*, 19 (3):368–375, 2003.

D. M. Rom. A sequentially rejective test procedure based on a modified Bonferroni inequality. *Biometrika*, 77:663–665, 1990.

J. P. Shaffer. Multiple hypothesis testing. *Annu. Rev. Psychol.*, 46:561–584, 1995.

R. J. Simes. An improved Bonferroni procedure for multiple tests of significance. *Biometrika*, 73:751–754, 1986.

J. D. Storey. The positive false discovery rate: A Bayesian interpretation and the q-value. Technical Report 2001-12, Department of Statistics, Stanford University, 2001. (In press, Annals of Statistics).

J. D. Storey. A direct approach to false discovery rates. *Journal of the Royal Statistical Society, Series B*, 64:479–498, 2002.

J. D. Storey, J. E. Taylor, and D. O. Siegmund. Strong control, conservative point estimation, and simultaneous conservative consistency of false discovery rates: A unified approach. *Journal of the Royal Statistical Society, Series B*, 2003. (In press).

J. D. Storey and R. Tibshirani. Estimating false discovery rates under dependence, with applications to DNA microarrays. Technical Report 2001-28, Department of Statistics, Stanford University, 2001.

J. D. Storey and R. Tibshirani. SAM thresholding and false discovery rates under dependence, with applications to DNA microarrays. Technical report, Department of Statistics, Stanford University, 2003.

V. Goss Tusher, R. Tibshirani, and G. Chu. Significance analysis of microarrays applied to transcriptional responses to ionizing radiation. *Proc. Natl. Acad. Sci.*, 98:5116–5121, 2001.

M. J. van der Laan, S. Dudoit, and K. S. Pollard. Multiple testing. Part II. Step-down procedures for control of the family-wise error rate. Technical Report 139, Division of Biostatistics, UC Berkeley, 2003. URL `www.bepress.com/ucbbiostat/paper139/`.

P. H. Westfall and S. S. Young. *Resampling-based multiple testing: Examples and methods for p-value adjustment*. John Wiley & Sons, 1993.

D. Yekutieli and Y. Benjamini. Resampling-based false discovery rate controlling multiple test procedures for correlated test statistics. *Journal of Statistical Planning and Inference*, 82:171–196, 1999.