Multiple Testing. Part II. Step-Down Procedures for Control of the Family-Wise Error Rate

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

This working paper is hosted by The Berkeley Electronic Press (bepress) and may not be commercially reproduced without the permission of the copyright holder.

<http://biostats.bepress.com/ucbbiostat/paper139>

Copyright ©2003 by the authors.

Multiple Testing. Part II. Step-Down Procedures for Control of the Family-Wise Error Rate

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

Abstract

The present article proposes two step-down multiple testing procedures for asymptotic control of the family-wise error rate (FWER): the first procedure is based on maxima of test statistics (step-down maxT), while the second relies on minima of unadjusted p-values (step-down minP). A key feature of our approach is the test statistics null distribution (rather than data generating null distribution) used to derive cut-offs (i.e., rejection regions) for these test statistics and the resulting adjusted p-values. For general null hypotheses, corresponding to submodels for the data generating distribution, we identify an asymptotic domination condition for a null distribution under which the step-down maxT and minP procedures asymptotically control the Type I error rate, for arbitrary data generating distributions, without the need for conditions such as subset pivotality. Inspired by this general characterization of a null distribution, we then propose as an explicit null distribution the asymptotic distribution of the vector of null-value shifted and scaled test statistics. Step-down procedures based on consistent estimators of the null distribution are shown to also provide asymptotic control of the Type I error rate. A general bootstrap algorithm is supplied to conveniently obtain consistent estimators of the null distribution.

Contents

1	Introduction	2
1.1	Multiple hypothesis testing framework	2
1.2	Outline	6
2	Step-down procedures for control of the family-wise error rate	9
2.1	Step-down procedure based on maxima of test statistics	9
2.1.1	Step-down maxT procedure	9
2.1.2	Asymptotic control of FWER	10
2.1.3	Explicit proposal for the test statistics null distribution	15
2.1.4	Adjusted p -values	17
2.2	Step-down procedure based on minima of unadjusted p -values	19
2.2.1	Step-down minP procedure	19
2.2.2	Asymptotic control of FWER	22
2.2.3	Explicit proposal for the test statistics null distribution	25
2.2.4	Adjusted p -values	27
3	Bootstrap-based step-down procedures for control of the family-wise error rate	29
3.1	Asymptotic control for consistent estimator of the null distribution	29
3.2	Bootstrap estimation of the null distribution	33

1 Introduction

1.1 Multiple hypothesis testing framework

The present article is concerned with *step-down* multiple testing procedures for controlling the *family-wise error rate* (i.e., the probability of at least one Type I error), when testing general null hypotheses defined in terms of sub-models for the data generating distribution. Our approach is based on a null distribution for the test statistics, rather than a data generating null distribution, and provides asymptotic control of the Type I error rate for general data generating distributions, without the need for conditions such as subset pivotality (Westfall and Young (1993), p. 42–43). The companion article (Dudoit et al., 2003b) gives a detailed introduction to our general approach to multiple testing and provides *single-step* multiple testing procedures for controlling Type I error rates defined as arbitrary parameters of the distribution of the number of Type I errors. The third article in this series proposes simple augmentations of FWER-controlling procedures which control the generalized family-wise error rate (i.e., the probability of at least $(k + 1)$ Type I errors, for some integer $k \geq 0$) and the proportion of false positives among the rejected hypotheses, under general data generating distributions, with arbitrary dependence structures among variables (van der Laan et al., 2003). We follow the framework described in the companion article on single-step procedures and refer the reader to Sections 1 and 2 of this earlier article for a detailed introduction (Dudoit et al., 2003b). The basic set-up and main definitions are recalled below for convenience.

As in Dudoit et al. (2003b), we adopt the following definitions for inverses of cumulative distribution functions (c.d.f.) and survivor functions. Let F denote a (non-decreasing and right-continuous) c.d.f. and let $\bar{F}$ denote the corresponding (non-increasing and right-continuous) survivor function, defined as $\bar{F} \equiv 1 - F$. For $\alpha \in [0, 1]$, define inverses as

$$F^{-1}(\alpha) \equiv \inf\{x : F(x) \geq \alpha\} \quad \text{and} \quad \bar{F}^{-1}(\alpha) \equiv \inf\{x : \bar{F}(x) \leq \alpha\}. \quad (1)$$

With these definitions, $\bar{F}^{-1}(\alpha) = F^{-1}(1 - \alpha)$.

Model. [Section 2.1.1 in Dudoit et al. (2003b)] Let $X_1, \dots, X_n$ be n independent and identically distributed (i.i.d.) random d -vectors, $X = (X(j) : j = 1, \dots, d) \sim P \in \mathcal{M}$, where the *data generating distribution* P is known to

be an element of a particular *statistical model* $\mathcal{M}$ (possibly non-parametric). For example, in cancer microarray studies, $(X_i(1), \dots, X_i(g))$ may denote a g -vector of gene expression measures and $(X_i(g+1), \dots, X_i(d))$ a $(d-g)$ -vector of biological and clinical outcomes for patient i , $i = 1, \dots, n$.

Null hypotheses. [Section 2.1.3 in Dudoit et al. (2003b)] In order to cover a broad class of testing problems, we define m null hypotheses in terms of a collection of *submodels*, $\mathcal{M}_j \subseteq \mathcal{M}$, $j = 1, \dots, m$, for the data generating distribution P . The m *null hypotheses* are defined as $H_{0j} \equiv \mathbf{I}(P \in \mathcal{M}_j)$ and the corresponding *alternative hypotheses* as $H_{1j} \equiv \mathbf{I}(P \notin \mathcal{M}_j)$. Thus, H_{0j} is true, i.e., $H_{0j} = 1$, if $P \in \mathcal{M}_j$ and false otherwise.

This general representation of null hypotheses includes the familiar case of single-parameter null hypotheses. In this setting, we consider an m -vector of *parameters*, $\mu = (\mu(j) : j = 1, \dots, m)$, defined as functions $\mu(j) = \mu_j(P) \in \mathbb{R}$ of the unknown data generating distribution P , and specify each null hypothesis in terms of one of these parameters. Parameters of interest include means, differences in means, correlations, and can refer to linear models, generalized linear models, survival models (e.g., Cox proportional hazards model), time-series models, dose-response models, etc. One distinguishes between two types of testing problems for single parameters.

$$\begin{array}{ll} \text{One-sided tests} & H_{0j} = \mathbf{I}(\mu(j) \leq \mu_0(j)) \\ \text{vs.} & H_{1j} = \mathbf{I}(\mu(j) > \mu_0(j)), \quad j = 1, \dots, m. \end{array}$$

$$\begin{array}{ll} \text{Two-sided tests} & H_{0j} = \mathbf{I}(\mu(j) = \mu_0(j)) \\ \text{vs.} & H_{1j} = \mathbf{I}(\mu(j) \neq \mu_0(j)), \quad j = 1, \dots, m. \end{array}$$

The hypothesized null-values, $\mu_0(j)$, are frequently zero (e.g., no difference in mean expression levels for gene j between two populations of patients).

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 $S_0^c = S_0^c(P) \equiv \{j : H_{0j} \text{ is false}\} = \{j : P \notin \mathcal{M}_j\}$ be the set of $m_1 = m - m_0$ false null hypotheses, i.e., true positives. The goal of a multiple testing procedure is to accurately estimate the set S_0 , and thus its complement S_0^c , while controlling probabilistically the number of false positives at a user-supplied level α .

Test statistics. [Section 2.1.4 in Dudoit et al. (2003b)] The decisions to reject or not the null hypotheses are based on an m -vector of *test statistics*, $T_n = (T_n(j) : j = 1, \dots, m)$, that are functions of the data, $X_1, \dots, X_n$. Denote the (finite sample) joint distribution of the test statistics T_n by $Q_n = Q_n(P)$. It is assumed that large values of $T_n(j)$ provide evidence against the null hypothesis H_{0j} . For two-sided tests, one can take absolute values of the test statistics.

Multiple testing procedures. [Section 2.2 in Dudoit et al. (2003b)] A *multiple testing procedure* (MTP) produces a set S_n of rejected hypotheses, that *estimates* S_0^c , the set of false null hypotheses,

$$S_n = S(T_n, Q_0, \alpha) \equiv \{j : H_{0j} \text{ is rejected}\} \subseteq \{1, \dots, m\}. \quad (2)$$

As indicated by the long notation $S(T_n, Q_0, \alpha)$, the set S_n (or $\hat{S}_0^c$) depends on: (i) the data, $X_1, \dots, X_n$, through the test statistics T_n ; (ii) a null distribution, Q_0 , for the test statistics, used to compute 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 this and the companion articles, can be represented as

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

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

Type I error rates. [Section 2.3 in Dudoit et al. (2003b)] 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, and a *false negative*, or *Type II error*, is committed when the test procedure fails to reject a false null hypothesis. Denote the number of Type I errors by $V_n = V(Q_0 | Q_n) \equiv |S(T_n, Q_0, \alpha) \cap S_0(P)| = |S_n \cap S_0|$, where the longer notation $V(Q_0 | Q_n)$ emphasizes the dependence of the distribution for the number of Type I errors on the null distribution Q_0 , used to derive cut-offs for the test statistics T_n , and on the true underlying distribution $Q_n = Q_n(P)$ for these test statistics (here, the subset S_0 is kept fixed at the truth $S_0(P)$ and the nominal level α of the test is also held fixed).

As in the companion article (Dudoit et al., 2003b), we consider error rates that are defined as functions of the distribution of the number of Type I errors, that is, can be represented as parameters $\theta(F_{V_n})$, where F_{V_n} is the discrete cumulative distribution function on $\{0, \dots, m\}$ for the number of Type I errors, V_n . Here, we focus on control of the *family-wise error rate* (FWER), or probability of at least one Type I error,

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

van der Laan et al. (2003) provide simple augmentations of FWER-controlling procedures that control the *generalized family-wise error rate* ($gFWER(k) \equiv 1 - F_{V_n}(k)$, for a user-supplied integer $k \geq 0$) and the *proportion of false positives among the rejected hypotheses* ($PFP(q) \equiv Pr(V_n/R_n > q)$, for a user-supplied $q \in (0, 1)$), under general data generating distributions P , with arbitrary dependence structures among variables.

Adjusted p -values. [Section 2.4 in Dudoit et al. (2003b)] Given any multiple testing procedure

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

based on cut-offs $c_j(\alpha) = c_j(T_n, Q_0, \alpha)$, the *adjusted p -value*, $\tilde{P}_{0n}(j) = \tilde{P}(j, T_n, Q_0)$, for null hypothesis H_{0j} , is defined as

$$\begin{aligned} \tilde{P}_{0n}(j) &\equiv \inf \{\alpha \in [0, 1] : j \in S(T_n, Q_0, \alpha)\} \\ &= \inf \{\alpha \in [0, 1] : c_j(T_n, Q_0, \alpha) < T_n(j)\}, \quad j = 1, \dots, m. \end{aligned} \quad (4)$$

That is, $\tilde{P}_{0n}(j)$ is the nominal level of the *entire* MTP (e.g., gFWER or FDR) at which H_{0j} would just be rejected, given T_n . For continuous null distributions Q_0 , $\tilde{P}_{0n}(j) = c_j^{-1}(T_n(j))$, where c_j^{-1} is the inverse of the monotone decreasing function $\alpha \rightarrow c_j(\alpha) = c_j(T_n, Q_0, \alpha)$. The particular mapping c_j , defining the cut-offs $c_j(T_n, Q_0, \alpha)$, will depend on the choice of MTP (e.g., single-step vs. stepwise, common cut-offs vs. common-quantile cut-offs).

In contrast, the *unadjusted p -value* (a.k.a. *marginal* or *raw p -value*), $P_{0n}(j) = P(T_n(j), Q_{0j})$, for the test of single null hypothesis H_{0j} , based on cut-offs $c_j(Q_{0j}, \alpha) = \bar{Q}_{0j}^{-1}(\alpha)$, involves only the marginal distribution Q_{0j} of the test statistic $T_n(j)$ for that hypothesis

$$P_{0n}(j) \equiv \inf \{\alpha \in [0, 1] : c_j(Q_{0j}, \alpha) < T_n(j)\}, \quad j = 1, \dots, m. \quad (5)$$

That is, $P_{0n}(j)$ is the nominal level of the *single* hypothesis testing procedure at which H_{0j} would just be rejected, given $T_n(j)$. For continuous marginal null distributions Q_{0j} , the unadjusted p -values are given by $P_{0n}(j) = c_j^{-1}(T_n(j)) = \bar{Q}_{0j}(T_n(j))$, where c_j^{-1} is the inverse of the monotone decreasing function $\alpha \rightarrow c_j(\alpha) = c_j(Q_{0j}, \alpha)$.

Stepwise procedures. [Section 2.5 in Dudoit et al. (2003b)] One usually distinguishes between two main classes of multiple testing procedures, single-step and stepwise procedures, depending on whether the cut-off vector $c = (c_j : j = 1, \dots, m)$ for the test statistics T_n is constant or random (given Q_0), i.e., is independent or not of these test statistics.

In *single-step* procedures, each hypothesis H_{0j} is evaluated using a critical value $c_j = c_j(Q_0, \alpha)$ that is independent of the results of the tests of other hypotheses and is not a function of the data $X_1, \dots, X_n$ (unless these data are used to estimate the null distribution Q_0 , as in Section 3).

Improvement in power, while preserving (asymptotic) Type I error rate control, may be achieved by *stepwise procedures*, in which rejection of a particular hypothesis depends on the outcome of the tests of other hypotheses. That is, the cut-offs $c_j = c_j(T_n, Q_0, \alpha)$ are allowed to depend on the data, $X_1, \dots, X_n$, via the test statistics T_n . In *step-down* procedures, the hypotheses corresponding to the *most significant* test statistics (i.e., largest absolute test statistics or smallest unadjusted p -values) are considered successively, with further tests depending on the outcome of earlier ones. As soon as one fails to reject a null hypothesis, no further hypotheses are rejected. In contrast, for *step-up* procedures, the hypotheses corresponding to the *least significant* test statistics are considered successively, again with further tests depending on the outcome of earlier ones. As soon as one hypothesis is rejected, all remaining more significant hypotheses are rejected.

1.2 Outline

Section 2 proposes two step-down multiple testing procedures for controlling the family-wise error rate (FWER), when testing general null hypotheses defined in terms of submodels for the data generating distribution. The first procedure relies on successive maxima of test statistics (*step-down maxT*, Procedure 1) and the second involves successive minima of unadjusted p -values (*step-down minP*, Procedure 2). We derive two main types of results concerning asymptotic control of the FWER by Procedures 1 and 2, under

a null distribution for the test statistics, rather than a data generating null distribution. The more general Theorems 1 and 4 prove that the step-down maxT and minP procedures provide asymptotic control of the FWER, under general asymptotic domination conditions for the null distribution and imply that gains in power from step-down procedures, relative to their single-step counterparts, do not come at the expense of Type I error control. By making additional asymptotic separation assumptions, Theorems 2 and 5 provide sharper control results. Theorem 3 proposes as an explicit null distribution the asymptotic distribution of the vector of null-value shifted and scaled test statistics. In Section 3, step-down maxT and minP procedures, based on a consistent estimator of the null distribution, are shown to also provide asymptotic control of the Type I error rate (Theorems 6 and 8). A general bootstrap procedure is supplied to conveniently obtain consistent estimators of the null distribution (Procedure 3). The proposed methods are evaluated by a simulation study and applied to gene expression microarray data in the fourth article of the series (Pollard et al., 2004). Software implementing the bootstrap single-step and step-down multiple testing procedures will be available in the R package `multtest`, released as part of the Bioconductor Project (www.bioconductor.org).

2 Step-down procedures for control of the family-wise error rate

2.1 Step-down procedure based on maxima of test statistics

2.1.1 Step-down maxT procedure

Procedure 1. Step-down maxT procedure for control of the FWER.

Let $T_n^\circ(j)$ be the ordered test statistics, $T_n^\circ(1) \geq \dots \geq T_n^\circ(m)$, and $O_n(j)$ the indices for these ordered statistics, so that $T_n^\circ(j) \equiv T_n(O_n(j))$, $j = 1, \dots, m$. For a level $\alpha \in (0, 1)$ test, given an m -variate null distribution Q_0 and random m -vector $Z = (Z(j) : j = 1, \dots, m) \sim Q_0$, define $(1 - \alpha)$ -quantiles, $c(A) = c(A, Q_0, \alpha) \in \mathbb{R}$, for the distributions of maxima, $\max_{j \in A} Z(j)$, of random variables $Z(j)$ over subsets $A \subseteq \{1, \dots, m\}$,

$$c(A, Q_0, \alpha) \equiv F_{A, Q_0}^{-1}(1 - \alpha) = \inf \{z : F_{A, Q_0}(z) \geq 1 - \alpha\}, \quad (6)$$

where $F_{A, Q_0}(z) \equiv \Pr_{Q_0}(\max_{j \in A} Z(j) \leq z)$ denotes the c.d.f. of $\max_{j \in A} Z(j)$ under $Z \sim Q_0$. Next, given the indices $O_n(j)$ for the ordered statistics $T_n^\circ(j)$, define $(1 - \alpha)$ -quantiles, $C_n(j)$, for subsets of the form $\overline{O}_n(j) \equiv \{O_n(j), \dots, O_n(m)\}$,

$$C_n(j) \equiv c(\overline{O}_n(j), Q_0, \alpha) = F_{\overline{O}_n(j), Q_0}^{-1}(1 - \alpha), \quad (7)$$

and step-down cut-offs

$$\begin{aligned} C_n^\circ(1) &\equiv C_n(1) \\ C_n^\circ(j) &\equiv \begin{cases} C_n(j), & \text{if } T_n^\circ(j-1) > C_n^\circ(j-1) \\ +\infty, & \text{otherwise} \end{cases}, \quad j = 2, \dots, m. \end{aligned} \quad (8)$$

The *step-down maxT* multiple testing procedure for controlling the FWER at level α is defined by the following rule: Reject null hypothesis $H_{0, O_n(j)}$, corresponding to the j th most significant test statistic $T_n^\circ(j) = T_n(O_n(j))$, if $T_n^\circ(j) > C_n^\circ(j)$, $j = 1, \dots, m$, that is,

$$S(T_n, Q_0, \alpha) \equiv \{O_n(j) : T_n^\circ(j) > C_n^\circ(j), j = 1, \dots, m\}. \quad (9)$$

Procedure 1 can be stated more compactly as

$$S(T_n, Q_0, \alpha) \equiv \{O_n(1), \dots, O_n(R_n)\},$$

where R_n , the number of rejected hypotheses, is defined as

$$R_n \equiv \max \left\{ j : \left(\sum_{l=1}^j I(T_n^\circ(l) > C_n(l)) \right) = j \right\}. \quad (10)$$

Note that the definition $C_n^\circ(j) = +\infty$, if $T_n^\circ(j-1) \leq C_n^\circ(j-1)$, ensures that the procedure is indeed step-down, that is, one can only reject a particular hypothesis provided all hypotheses with more significant (i.e., larger) test statistics were rejected beforehand. In addition, the cut-offs $C_n(j)$ used in the rejection rule are random variables that depend on the data via the ranks of the test statistics T_n (i.e., via the random subsets $\overline{O}_n(j)$), again reflecting the stepwise nature of the procedure. This is in contrast to the constant cut-offs used in single-step Procedures 1 and 2 of the companion article (Dudoit et al., 2003b).

Procedure 1, based on successive maxima of test statistics, is a step-down analogue of the *single-step maxT* procedure that arises as a special case of *single-step common-cut-off* Procedure 2 of Dudoit et al. (2003b). Similar step-down maxT procedures are discussed in Dudoit et al. (2003a) and Westfall and Young (1993), Algorithm 4.1, p. 116–117, with an important distinction in the choice of the null distribution Q_0 used to derive the quantiles $C_n(j)$ (and the resulting adjusted p -values in Section 2.1.4).

2.1.2 Asymptotic control of FWER

In order to establish asymptotic control of the FWER by Procedure 1, we rely on one or both of the following two assumptions concerning the joint distribution $Q_n = Q_n(P)$ of the test statistics T_n and the null distribution Q_0 .

Assumption AT1 [Asymptotic null domination] There exists an m -variate null distribution $Q_0 = Q_0(P)$ so that

$$\limsup_{n \rightarrow \infty} Pr_{Q_n} \left(\max_{j \in S_0} T_n(j) > x \right) \leq Pr_{Q_0} \left(\max_{j \in S_0} Z(j) > x \right) \text{ for all } x, \quad (11)$$

where T_n and Z are random m -vectors with $T_n \sim Q_n = Q_n(P)$ and $Z \sim Q_0 = Q_0(P)$.

Assumption AT2 [Asymptotic separation of true and false null hypotheses] Let M_1 be a possibly degenerate (e.g., $+\infty$) maximal value, so that $Pr_{Q_n}(\max_j T_n(j) < M_1) = 1$, for all n . Assume that for all $M < M_1$,

$$\lim_{n \rightarrow \infty} Pr_{Q_n} \left(\min_{j \in S_0^c} T_n(j) \geq M \right) = 1 \quad (12)$$

and

$$\lim_{M \uparrow M_1} \lim_{n \rightarrow \infty} Pr_{Q_n} \left(\max_{j \in S_0} T_n(j) \geq M \right) = 0. \quad (13)$$

In addition, for $\alpha \in (0, 1)$ and $Z = (Z(j) : j = 1, \dots, m) \sim Q_0$, the distributions of maxima, $\max_{j \in A} Z(j)$, of random variables $Z(j)$ over subsets $A \subseteq \{1, \dots, m\}$, are assumed to have $(1 - \alpha)$ -quantiles bounded by M_1 , that is,

$$\max_{A \subseteq \{1, \dots, m\}} c(A, Q_0, \alpha) < M_1, \quad (14)$$

where the quantiles $c(A, Q_0, \alpha)$ are defined as in Procedure 1.

Note that Assumption AT1 follows from the *asymptotic null domination condition* AQ0 in Theorem 1 of the companion article on single-step procedures (Dudoit et al., 2003b). Condition AQ0 was stated there to asymptotically control general Type I error rates of the form $\theta(F_{V_n})$ and it is sufficient, but not necessary, for control of the FWER. Assumption AT1 can therefore be viewed as an FWER-specific asymptotic null domination condition, where $\theta(F_{V_n}) = 1 - F_{V_n}(0)$. Specific guidelines for constructing a null distribution Q_0 that satisfies Assumption AT1 are given in Theorem 3. Also note that conditions (12) and (13) in Assumption AT2 only require that $T_n = (T_n(j) : j = 1, \dots, m)$ represent a sensible set of test statistics, that separate into two groups as $n \rightarrow \infty$, depending on the truth or falsity of the null hypotheses, where the largest m_1 test statistics correspond to the m_1 false null hypotheses.

We derive two main results concerning asymptotic control of the FWER by Procedure 1. The more general Theorem 1 proves that Procedure 1 provides asymptotic control of the FWER under asymptotic null domination

Assumption AT1 only. Since step-down cut-offs are always less than or equal to the corresponding single-step cut-offs, this result shows that the gain in power from the step-down maxT procedure, relative to the single-step maxT procedure, does not come at the expense of failure of Type I error control. By making the additional asymptotic separation Assumption AT2, Theorem 2 provides a sharper result. In particular, consistent identification of the set S_0^c of false null hypotheses as in Assumption AT2, leads to exact asymptotic control of the FWER, when condition (11) in Assumption AT1 holds with equality. However, asymptotic separation of the test statistics for the true and false null hypotheses does not hold at local alternatives.

Theorem 1 [Asymptotic control of FWER for step-down maxT Procedure 1, under Assumption AT1] *Suppose Assumption AT1 of asymptotic null domination is satisfied by the distribution $Q_n = Q_n(P)$ of the test statistics $T_n = (T_n(j) : j = 1, \dots, m)$ and the null distribution Q_0 . Denote the number of Type I errors for Procedure 1 by*

$$V_n \equiv \sum_{j=1}^m \mathbf{I}(T_n^o(j) > C_n^o(j), O_n(j) \in S_0).$$

Then, Procedure 1 provides asymptotic control of the family-wise error rate at level α , that is,

$$\limsup_{n \rightarrow \infty} \Pr(V_n > 0) \leq \alpha.$$

Proof of Theorem 1. Let $J_n \equiv \min\{j : O_n(j) \in S_0\}$, that is, $O_n(J_n)$ is the index of the *true* null hypothesis with the largest test statistic. Thus, by definition of J_n , $T_n^o(J_n) = \max_{j \in S_0} T_n(j)$ and $\{O_n(1), \dots, O_n(J_n - 1)\} = \overline{O_n(J_n)}^c \subseteq S_0^c$. It then follows that

$$\begin{aligned} \Pr(V_n > 0) &= \Pr(O_n(J_n) \in S_n) \\ &\leq \Pr(T_n^o(J_n) > c(\overline{O_n(J_n)}, Q_0, \alpha)) \\ &= \Pr\left(\max_{j \in S_0} T_n(j) > c(\overline{O_n(J_n)}, Q_0, \alpha)\right) \\ &\leq \Pr\left(\max_{j \in S_0} T_n(j) > c(S_0, Q_0, \alpha)\right), \end{aligned}$$

where the first inequality follows from the step-down property and the last inequality follows from the fact that $S_0 \subseteq \overline{O_n(J_n)}$ implies $c(S_0, Q_0, \alpha) \leq$

$c(\overline{O}_n(J_n), Q_0, \alpha)$. Finally, under Assumption AT1 and for $Z \sim Q_0$,

$$\begin{aligned} \limsup_{n \rightarrow \infty} Pr(V_n > 0) &\leq \limsup_{n \rightarrow \infty} Pr\left(\max_{j \in S_0} T_n(j) > c(S_0, Q_0, \alpha)\right) \\ &\leq Pr\left(\max_{j \in S_0} Z(j) > c(S_0, Q_0, \alpha)\right) \\ &\leq \alpha, \end{aligned}$$

which completes the proof. □

Theorem 2 [Asymptotic control of FWER for step-down maxT Procedure 1, under Assumptions AT1 and AT2] *Suppose Assumptions AT1 and AT2 hold, specifically, conditions (11), (12), (13), and (14), are satisfied by the distribution $Q_n = Q_n(P)$ of the test statistics $T_n = (T_n(j) : j = 1, \dots, m)$ and the null distribution Q_0 . Denote the number of Type I errors for Procedure 1 by*

$$V_n \equiv \sum_{j=1}^m I(T_n^\circ(j) > C_n^\circ(j), O_n(j) \in S_0).$$

Then, Procedure 1 provides asymptotic control of the family-wise error rate at level α , that is,

$$\limsup_{n \rightarrow \infty} Pr(V_n > 0) \leq \alpha.$$

If condition (11) in Assumption AT1 holds with equality and Q_0 is continuous (so that $\max_{j \in S_0} Z(j)$ is a continuous random variable for $Z \sim Q_0$), then asymptotic control is exact

$$\lim_{n \rightarrow \infty} Pr(V_n > 0) = \alpha.$$

Proof of Theorem 2. Procedure 1 can be stated equivalently in terms of statistics $T_n^\star(j)$, as follows. Let

$$\begin{aligned} T_n^\star(1) &\equiv T_n^\circ(1) \\ T_n^\star(j) &\equiv \begin{cases} T_n^\circ(j), & \text{if } T_n^\star(j-1) > C_n(j-1) \\ -\infty, & \text{otherwise} \end{cases}, \quad j = 2, \dots, m, \end{aligned} \tag{15}$$

and reject $H_{0,O_n(j)}$ if $T_n^*(j) > C_n(j)$, $j = 1, \dots, m$. We first state the main ideas of the proof. Note that, from asymptotic separation Assumption AT2, with probability one in the limit, the first $m_1 = |S_0^c|$ rejected hypotheses correspond to the m_1 false null hypotheses (see argument with indicator B_n , below). Thus, no Type I errors are committed for these first m_1 rejections and one can focus on the m_0 least significant statistics, $T_n^\circ(j)$, $j = m_1 + 1, \dots, m$, which now correspond to the test statistics for the true null hypotheses, $(T_n(j) : j \in S_0)$. By definition of the step-down procedure, a Type I error is then committed if and only if $T_n^*(m_1 + 1) = \max_{j \in S_0} T_n(j) > C_n(m_1 + 1) = c(S_0)$. Thus, one needs to control $Pr(\max_{j \in S_0} T_n(j) > c(S_0))$, which the procedure indeed asymptotically controls at level α , conditional on the event that the first m_1 rejections correspond exactly with rejecting the true positives S_0^c . Details of the proof are given next.

Define Bernoulli random variables

$$B_n \equiv \mathbb{I}(\{O_n(1), \dots, O_n(m_1)\} = S_0^c, T_n^*(1) > C_n(1), \dots, T_n^*(m_1) > C_n(m_1)). \quad (16)$$

Under Assumption AT2, $Pr(B_n = 1) \rightarrow 1$ as $n \rightarrow \infty$. Then, the FWER for Procedure 1 is given by

$$\begin{aligned} Pr(V_n > 0) &= Pr(\cup_{j=1}^m \{T_n^*(j) > C_n(j), O_n(j) \in S_0\}) \\ &= Pr(\cup_{j=1}^m \{T_n^*(j) > C_n(j), O_n(j) \in S_0\} \mid B_n = 1) + o(1) \\ &= Pr(\cup_{j=m_1+1}^m \{T_n^*(j) > C_n(j)\} \mid B_n = 1) + o(1) \\ &= E \left[Pr(\cup_{j=m_1+1}^m \{T_n^*(j) > C_n(j)\} \mid \overline{O}_n(m_1 + 1), B_n = 1) \mid B_n = 1 \right] + o(1) \\ &= E \left[Pr(\cup_{j=m_1+1}^m \{T_n^*(j) > C_n(j)\} \mid \mathbb{I}(T_n^\circ(m_1 + 1) > c(S_0)) * B_n = 1, \overline{O}_n(m_1 + 1)) \right. \\ &\quad \times \left. Pr(T_n^\circ(m_1 + 1) > c(S_0) \mid \overline{O}_n(m_1 + 1), B_n = 1) \mid B_n = 1 \right] + o(1), \end{aligned}$$

where $B_n = 1$ implies that $\overline{O}_n(m_1 + 1) = S_0$ and hence $C_n(m_1 + 1) = c(S_0)$. The last equality follows by noting that, given $B_n = 1$, then $T_n^*(m_1 + 1) = T_n^\circ(m_1 + 1) = T_n(O_n(m_1 + 1))$, and, if $T_n^*(m_1 + 1) \leq C_n(m_1 + 1)$, then $T_n^*(j) = -\infty$ for $j = m_1 + 2, \dots, m$. Now, note that the first probability

within the conditional expectation equals one, so that

$$\begin{aligned}
Pr(V_n > 0) &= \\
&= E \left[Pr(T_n^\circ(m_1 + 1) > c(S_0) \mid \bar{O}_n(m_1 + 1), B_n = 1) \mid B_n = 1 \right] + o(1) \\
&= E \left[Pr \left(\max_{j \in S_0} T_n(j) > c(S_0) \mid \bar{O}_n(m_1 + 1), B_n = 1 \right) \mid B_n = 1 \right] + o(1) \\
&= Pr \left(\max_{j \in S_0} T_n(j) > c(S_0) \mid B_n = 1 \right) + o(1) \\
&= Pr \left(\max_{j \in S_0} T_n(j) > c(S_0) \right) + o(1),
\end{aligned}$$

where we again use the fact that $Pr(B_n = 1) \rightarrow 1$.

Finally, under Assumption AT1 and for $Z \sim Q_0$,

$$\limsup_{n \rightarrow \infty} Pr \left(\max_{j \in S_0} T_n(j) > c(S_0) \right) \leq Pr \left(\max_{j \in S_0} Z(j) > c(S_0) \right) \leq \alpha.$$

If condition (11) in Assumption AT1 holds with equality (i.e., $\lim_n Pr(\max_{j \in S_0} T_n(j) > x) = Pr(\max_{j \in S_0} Z(j) > x)$) and the null distribution Q_0 is continuous, so that quantiles $c(A) = c(A, Q_0, \alpha)$ provide exact α survival probabilities, then we have exact asymptotic control at level α

$$\lim_{n \rightarrow \infty} Pr \left(\max_{j \in S_0} T_n(j) > c(S_0) \right) = Pr \left(\max_{j \in S_0} Z(j) > c(S_0) \right) = \alpha.$$

□

2.1.3 Explicit proposal for the test statistics null distribution

One can make the following explicit proposal for a null distribution Q_0 that satisfies asymptotic null domination Assumption AT1.

Theorem 3 [General construction for null distribution Q_0] *Suppose there exists known m -vectors $\lambda_0 \in \mathbb{R}^m$ and $\tau_0 \in \mathbb{R}^{+m}$ of null-values, so that*

$$\begin{aligned}
\limsup_{n \rightarrow \infty} E[T_n(j)] &\leq \lambda_0(j) & \text{and} & & (17) \\
\limsup_{n \rightarrow \infty} Var[T_n(j)] &\leq \tau_0(j), & \text{for } j \in S_0. &
\end{aligned}$$

Let

$$\nu_{0n}(j) \equiv \sqrt{\min \left(1, \frac{\tau_0(j)}{\text{Var}[T_n(j)]} \right)} \quad (18)$$

and define an m -vector Z_n by

$$Z_n(j) \equiv \nu_{0n}(j) \left(T_n(j) + \lambda_0(j) - E[T_n(j)] \right), \quad j = 1, \dots, m. \quad (19)$$

Suppose that

$$Z_n \xrightarrow{\mathcal{L}} Z \sim Q_0(P). \quad (20)$$

Then, for this choice of null distribution $Q_0 = Q_0(P)$, and for all x ,

$$\limsup_{n \rightarrow \infty} Pr_{Q_n} \left(\max_{j \in S_0} T_n(j) > x \right) \leq Pr_{Q_0} \left(\max_{j \in S_0} Z(j) > x \right), \quad (21)$$

so that asymptotic null domination condition (11) in Assumption AT1 holds. In particular, if for all $j \in S_0$, $\lim_n E[T_n(j)] = \lambda_0(j)$, then (21) holds with equality.

Proof of Theorem 3. Define an intermediate random vector $(\tilde{Z}_n(j) : j \in S_0)$, for the true null hypotheses, by

$$\tilde{Z}_n(j) \equiv T_n(j) + \max(0, \lambda_0(j) - E[T_n(j)]), \quad j \in S_0. \quad (22)$$

Then, $T_n(j) \leq \tilde{Z}_n(j)$. In addition, since $\limsup_n E[T_n(j)] \leq \lambda_0(j)$ and $\limsup_n \text{Var}[T_n(j)] \leq \tau_0(j)$ for $j \in S_0$ (and thus $\lim_n \nu_{0n}(j) = 1$), it follows that $(\tilde{Z}_n(j) : j \in S_0)$ and $(Z_n(j) : j \in S_0)$ have the same limit distribution

$$(\tilde{Z}_n(j) : j \in S_0) \xrightarrow{\mathcal{L}} (Z(j) : j \in S_0) \sim Q_{0,S_0}.$$

Thus, by the Continuous Mapping Theorem,

$$\begin{aligned} \limsup_{n \rightarrow \infty} Pr \left(\max_{j \in S_0} T_n(j) > x \right) &\leq \limsup_{n \rightarrow \infty} Pr \left(\max_{j \in S_0} \tilde{Z}_n(j) > x \right) \\ &= Pr \left(\max_{j \in S_0} Z(j) > x \right) \text{ for all } x. \end{aligned}$$

In particular, if (17) holds with equality, then (21) also holds with equality.

□

The null distribution Q_0 proposed in Theorem 3 for step-down procedures is the same as the general null distribution proposed for single-step procedures in Theorem 2 of the companion article (Dudoit et al., 2003b). The reader is referred to this earlier article for motivation for the construction of the null distribution Q_0 and a detailed discussion of its properties (Sections 2.6, 3.2, and 5). In particular, Section 5 provides null-values $\lambda_0(j)$ and $\tau_0(j)$ for a broad range of testing problems and also discusses null distributions for specific choices of test statistics. In many testing problems of interest, Q_0 is continuous. For instance, for the test of single-parameter null hypotheses using t -statistics, Q_0 is an m -variate Gaussian distribution with mean vector zero (Section 5.1).

In practice, one can estimate the null distribution Q_0 using a bootstrap procedure, as discussed in detail in Section 3. For B bootstrap samples, one has an $m \times B$ matrix of test statistics, $\mathbf{T} = (T_n^b(j))$, with rows corresponding to the m hypotheses and columns to the B bootstrap samples. The expected values, $E[T_n(j)]$, and variances, $\text{Var}[T_n(j)]$, are estimated by simply taking row means and variances of the matrix $\mathbf{T}$. The matrix of test statistics $\mathbf{T}$ can then be row-shifted and scaled using the supplied null-values $\lambda_0(j)$ and $\tau_0(j)$, to produce an $m \times B$ matrix $\mathbf{Z} = (Z_n^b(j))$. The null distribution Q_0 is estimated by the empirical distribution of the columns of matrix $\mathbf{Z}$.

2.1.4 Adjusted p -values

Rather than simply reporting rejection or not of a subset of null hypotheses at a prespecified level α , one can report *adjusted p -values* for step-down Procedure 1, computed under the assumed null distribution Q_0 for the test statistics T_n (for a more detailed discussion of adjusted p -values, consult Section 2.4 of the companion article, Dudoit et al. (2003b)). While the definition of adjusted p -value in equation (4) of Section 1 holds for general null distributions, in this section, we consider for simplicity a null distribution Q_0 with continuous and strictly monotone marginal c.d.f.'s, Q_{0j} , and survivor functions, $\bar{Q}_{0j} = 1 - Q_{0j}$, $j = 1, \dots, m$.

Result 1 [Adjusted p -values for step-down maxT Procedure 1] *The adjusted p -values for step-down maxT Procedure 1, based on a null distribution Q_0 with continuous and strictly monotone marginal distributions, are*

given by

$$\tilde{P}_{0n}(O_n(j)) = \max_{k=1, \dots, j} \left\{ Pr_{Q_0} \left(\max_{l \in \{O_n(k), \dots, O_n(m)\}} Z(l) \geq T_n(O_n(k)) \right) \right\}, \quad (23)$$

where $Z = (Z(j) : j = 1, \dots, m) \sim Q_0$ and $H_{0, O_n(j)}$ is the null hypothesis corresponding to the j th most significant test statistic $T_n^o(j) = T_n(O_n(j))$, that is, the indices $O_n(j)$ are defined such that $T_n(O_n(1)) \geq \dots \geq T_n(O_n(m))$. Step-down Procedure 1 for controlling the FWER at level α can then be stated equivalently as

$$S(T_n, Q_0, \alpha) = \{O_n(j) : \tilde{P}_{0n}(O_n(j)) \leq \alpha, j = 1, \dots, m\}.$$

Note that the adjusted p -values are conditional on the observed test statistics $T_n(j)$ and their ranks. In addition, taking successive maxima of the probabilities in equation (23) enforces the step-down property via monotonicity of the adjusted p -values, $\tilde{P}_{0n}(O_n(1)) \leq \dots \leq \tilde{P}_{0n}(O_n(m))$.

Proof of Result 1. As in Procedure 1, let $F_{A, Q_0}(z) \equiv Pr_{Q_0}(\max_{j \in A} Z(j) \leq z)$ denote the c.d.f. of $\max_{j \in A} Z(j)$ for $Z \sim Q_0$, $\bar{O}_n(j) \equiv \{O_n(j), \dots, O_n(m)\}$, and $C_n(j) = F_{\bar{O}_n(j), Q_0}^{-1}(1 - \alpha)$. Then,

$$\begin{aligned} \tilde{P}_{0n}(O_n(j)) &= \inf \left\{ \alpha \in [0, 1] : \sum_{k=1}^j I(T_n(O_n(k)) > C_n(k)) = j \right\} \\ &= \inf \{ \alpha \in [0, 1] : T_n(O_n(k)) > C_n(k), \forall k = 1, \dots, j \} \\ &= \max_{k=1, \dots, j} \inf \{ \alpha \in [0, 1] : T_n(O_n(k)) > C_n(k) \} \\ &= \max_{k=1, \dots, j} \inf \left\{ \alpha \in [0, 1] : T_n(O_n(k)) > F_{\bar{O}_n(k), Q_0}^{-1}(1 - \alpha) \right\} \\ &= \max_{k=1, \dots, j} \inf \left\{ \alpha \in [0, 1] : \bar{F}_{\bar{O}_n(k), Q_0}(T_n(O_n(k))) < \alpha \right\} \quad (\star) \\ &= \max_{k=1, \dots, j} \bar{F}_{\bar{O}_n(k), Q_0}(T_n(O_n(k))) \\ &= \max_{k=1, \dots, j} \left\{ Pr_{Q_0} \left(\max_{l \in \{O_n(k), \dots, O_n(m)\}} Z(l) > T_n(O_n(k)) \right) \right\}, \end{aligned}$$

where in $(\star)$ we use the fact that, for a c.d.f. F and corresponding survivor function $\bar{F}$, $\bar{F}^{-1}(\alpha) = F^{-1}(1 - \alpha)$. □

2.2 Step-down procedure based on minima of unadjusted p -values

2.2.1 Step-down minP procedure

One can also prove asymptotic control of the FWER for an analogue of Procedure 1, where maxima of test statistics, $T_n(j)$, are replaced by minima of unadjusted p -values, $P_{0n}(j)$, also computed under the null distribution Q_0 . Procedure 2, below, is a step-down analogue of the *single-step minP* procedure that arises as a special case of *single-step common-quantile* Procedure 1 of Dudoit et al. (2003b). Similar step-down minP procedures are discussed in Dudoit et al. (2003a) and Westfall and Young (1993), Algorithm 2.8, p. 66–67, with an important distinction in the choice of the null distribution Q_0 used to derive the quantiles $C_n(j)$ (and the resulting adjusted p -values in Section 2.2.4).

Note that procedures based on maxima of test statistics (maxT) and minima of unadjusted p -values (minP) are equivalent when the test statistics $T_n(j)$, $j = 1, \dots, m$, are identically distributed under Q_0 , i.e., when the marginal distributions Q_{0j} do not depend on j : in this case, the significance rankings based on test statistics $T_n(j)$ and marginal p -values $P_{0n}(j)$ coincide. In general, however, the two types of procedures produce different results, and considerations of balance, power, and computational feasibility should dictate the choice between the two approaches (Dudoit et al., 2003a,b; Ge et al., 2003). Also note that while nominal p -values computed from a standard normal or other distribution may not be correct, a step-down procedure based on minima of such transformed test statistics nonetheless provides asymptotic control of the FWER (e.g., $P_n(j) = \bar{\Phi}(T_n(j))$, where $\bar{\Phi}$ is the standard normal survivor function). That is, these p -values can be viewed as just another type of test statistic and one can apply Procedure 1 to $T_n(j) = -P_n(j)$ and appeal to Theorems 1, 2, and 3 for FWER control.

Here, however, we propose a step-down multiple testing procedure where unadjusted p -values are also defined in terms of the null distribution Q_0 . We therefore have a more specific procedure and assumptions for proving asymptotic control of the family-wise error rate than in Section 2.1. Specifically, define *unadjusted p -values* as $P_{0n}(j) \equiv \bar{Q}_{0j}(T_n(j))$, where $\bar{Q}_{0j} = 1 - Q_{0j}$ denote the marginal survivor functions corresponding to Q_0 , $j = 1, \dots, m$. Asymptotic Type I error control by step-down minP Procedure 2 relies on Assumptions AP1 and AP2, below; guidelines for constructing the null dis-

tribution Q_0 are given in Lemma 1 and are as in Theorem 3 with a few additional requirements.

Procedure 2. Step-down minP procedure for control of the FWER.

Given an m -variate null distribution Q_0 , with marginal c.d.f.'s Q_{0j} and survivor functions $\bar{Q}_{0j} = 1 - Q_{0j}$, $j = 1, \dots, m$, define unadjusted p -values

$$P_{0n}(j) \equiv \bar{Q}_{0j}(T_n(j)) \quad (24)$$

and

$$P_0(j) \equiv \bar{Q}_{0j}(Z(j)), \quad (25)$$

for random m -vectors $T_n = (T_n(j) : j = 1, \dots, m) \sim Q_n = Q_n(P)$ and $Z = (Z(j) : j = 1, \dots, m) \sim Q_0$. Let $P_{0n}^\circ(j)$ denote ordered unadjusted p -values, $P_{0n}^\circ(1) \leq \dots \leq P_{0n}^\circ(m)$, and $O_n(j)$ the indices for these ordered statistics, so that $P_{0n}^\circ(j) \equiv P_{0n}(O_n(j))$, $j = 1, \dots, m$. For a level $\alpha \in (0, 1)$ test, define α -quantiles, $c(A) = c(A, Q_0, \alpha) \in [0, 1]$, for the distributions of minima, $\min_{j \in A} P_0(j)$, of unadjusted p -values $P_0(j)$ over subsets $A \subseteq \{1, \dots, m\}$,

$$c(A, Q_0, \alpha) \equiv F_{A, Q_0}^{-1}(\alpha) = \inf \{z : F_{A, Q_0}(z) \geq \alpha\}, \quad (26)$$

where $F_{A, Q_0}(z) \equiv \Pr_{Q_0}(\min_{j \in A} P_0(j) \leq z)$ denotes the c.d.f. of $\min_{j \in A} P_0(j)$ for $Z \sim Q_0$. Next, given the indices $O_n(j)$ for the ordered unadjusted p -values $P_{0n}^\circ(j)$, define α -quantiles, $C_n(j)$, for subsets of the form $\bar{O}_n(j) \equiv \{O_n(j), \dots, O_n(m)\}$,

$$C_n(j) \equiv c(\bar{O}_n(j), Q_0, \alpha) = F_{\bar{O}_n(j), Q_0}^{-1}(\alpha), \quad (27)$$

and step-down cut-offs

$$\begin{aligned} C_n^\circ(1) &\equiv C_n(1) \\ C_n^\circ(j) &\equiv \begin{cases} C_n(j), & \text{if } P_{0n}^\circ(j-1) < C_n^\circ(j-1) \\ 0, & \text{otherwise} \end{cases}, \quad j = 2, \dots, m. \end{aligned} \quad (28)$$

The *step-down minP* multiple testing procedure for controlling the FWER at level α is defined by the following rule: Reject null hypothesis $H_{0, O_n(j)}$, corresponding to the j th most significant unadjusted p -value $P_{0n}^\circ(j) = P_{0n}(O_n(j)) = \bar{Q}_{0, O_n(j)}(T_n(O_n(j)))$, if $P_{0n}^\circ(j) < C_n^\circ(j)$, $j = 1, \dots, m$, that is,

$$S(T_n, Q_0, \alpha) \equiv \{O_n(j) : P_{0n}^\circ(j) < C_n^\circ(j), j = 1, \dots, m\}. \quad (29)$$

As for Procedure 1, Procedure 2 can also be stated more compactly as

$$S(T_n, Q_0, \alpha) = \{O_n(1), \dots, O_n(R_n)\}, \quad (30)$$

where R_n , the number of rejected hypotheses, is defined as

$$R_n \equiv \max \left\{ j : \left(\sum_{l=1}^j \mathbf{I}(P_{0n}^\circ(l) < C_n(l)) \right) = j \right\}. \quad (31)$$

The definition $C_n^\circ(j) = 0$, if $P_{0n}^\circ(j-1) \geq C_n^\circ(j-1)$, ensures that the procedure is indeed step-down, that is, one can only reject a particular hypothesis provided all hypotheses with more significant (i.e., smaller) unadjusted p -values were rejected beforehand.

Note that for a null distribution Q_0 with continuous margins, the unadjusted p -values $P_0(j) = \bar{Q}_{0j}(Z(j))$ have $U(0, 1)$ marginal distributions. However, the $P_0(j)$ are not independent, therefore, the quantiles $C_n(j)$ cannot be obtained trivially from the $Beta(1, m - j + 1)$ distribution. A key feature of Theorem 3 is that it provides a null distribution Q_0 for multiple testing procedures that take into account the *joint* distribution of the test statistics, i.e., the correlation structure of the null distribution Q_0 is implied by the correlation structure of the test statistics T_n , via the null-value shifted and scaled statistics Z_n .

2.2.2 Asymptotic control of FWER

As with step-down maxT Procedure 1, we prove two main theorems concerning asymptotic control of the FWER by Procedure 2, under the following p -value analogues of Assumptions AT1 and AT2. The more general result (Theorem 4) is proved under only asymptotic null domination Assumption AP1, and the sharper result (Theorem 5) is proved under both Assumptions AP1 and AP2. Guidelines for constructing a null distribution Q_0 that satisfies Assumptions AP1 and AP2 are as in Theorem 3, with a few additional conditions stated in Lemma 1 and 2, below.

Assumption AP1 [Asymptotic null domination] There exists an m -variate null distribution $Q_0 = Q_0(P)$ so that

$$\limsup_{n \rightarrow \infty} Pr_{Q_n} \left(\min_{j \in S_0} P_{0n}(j) < x \right) \leq Pr_{Q_0} \left(\min_{j \in S_0} P_0(j) < x \right) \text{ for all } x, \quad (32)$$

where $P_{0n}(j) = \bar{Q}_{0j}(T_n(j))$ and $P_0(j) = \bar{Q}_{0j}(Z(j))$ are unadjusted p -values defined for random m -vectors $T_n \sim Q_n = Q_n(P)$ and $Z \sim Q_0 = Q_0(P)$, respectively, and $\bar{Q}_{0j}$ denote the marginal survivor functions corresponding to the null distribution Q_0 , $j = 1, \dots, m$.

Note that, like Assumption AT1 for the step-down maxT procedure, Assumption AP1 follows from the *asymptotic null domination condition* AQ0 in Theorem 1 of the companion article on single-step procedures (Dudoit et al., 2003b). It is a weaker form of null domination, that is specific to FWER-controlling procedures based on p -values.

Assumption AP2 [Asymptotic separation of true and false null hypotheses] Let $P_{0n}(j) = \bar{Q}_{0j}(T_n(j))$ and $P_0(j) = \bar{Q}_{0j}(Z(j))$ denote unadjusted p -values defined for random m -vectors $T_n \sim Q_n = Q_n(P)$ and $Z \sim Q_0$, respectively, where $\bar{Q}_{0j}$ denote the marginal survivor functions corresponding to the null distribution Q_0 , $j = 1, \dots, m$. For each $\epsilon > 0$, assume that

$$\lim_{n \rightarrow \infty} Pr_{Q_n} \left(\max_{j \in S_0^c} P_{0n}(j) \leq \epsilon \right) = 1 \quad (33)$$

and

$$\lim_{\epsilon \downarrow 0} \lim_{n \rightarrow \infty} Pr_{Q_n} \left(\min_{j \in S_0} P_{0n}(j) \leq \epsilon \right) = 0. \quad (34)$$

In addition, for $\alpha \in (0, 1)$ and $Z = (Z(j) : j = 1, \dots, m) \sim Q_0$, the distributions of minima, $\min_{j \in A} P_0(j)$, of unadjusted p -values $P_0(j) = \bar{Q}_{0j}(Z(j))$ over subsets $A \subseteq \{1, \dots, m\}$, are assumed to have positive α -quantiles, that is,

$$\min_{A \subseteq \{1, \dots, m\}} c(A, Q_0, \alpha) > 0, \quad (35)$$

where the quantiles $c(A, Q_0, \alpha)$ are defined as in Procedure 2.

Theorem 4 [Asymptotic control of FWER for step-down minP Procedure 2, under Assumption AP1] Suppose Assumption AP1 of asymptotic null domination is satisfied by the unadjusted p -values, $P_{0n} = (P_{0n}(j) = \bar{Q}_{0j}(T_n(j)) : j = 1, \dots, m)$, i.e., by the distribution $Q_n = Q_n(P)$ of the test statistics $T_n = (T_n(j) : j = 1, \dots, m)$ and by the null distribution Q_0 . Denote

the number of Type I errors for Procedure 2 by

$$V_n \equiv \sum_{j=1}^m \mathbf{I}(P_{0n}^\circ(j) < C_n^\circ(j), O_n(j) \in S_0).$$

Then, Procedure 2 provides asymptotic control of the family-wise error rate at level α , that is,

$$\limsup_{n \rightarrow \infty} \Pr(V_n > 0) \leq \alpha.$$

Proof of Theorem 4. The proof follows that of Theorem 1, with unadjusted p -values $P_{0n}(j)$ replacing test statistics $T_n(j)$.

□

Theorem 5 [Asymptotic control of FWER for step-down minP Procedure 2, under Assumptions AP1 and AP2] Suppose Assumptions AP1 and AP2 hold, specifically, conditions (32), (33), (34), and (35), are satisfied by the unadjusted p -values, $P_{0n} = (P_{0n}(j) = \bar{Q}_{0j}(T_n(j)) : j = 1, \dots, m)$, i.e., by the distribution $Q_n = Q_n(P)$ of the test statistics $T_n = (T_n(j) : j = 1, \dots, m)$ and by the null distribution Q_0 . Denote the number of Type I errors for Procedure 2 by

$$V_n \equiv \sum_{j=1}^m \mathbf{I}(P_{0n}^\circ(j) < C_n^\circ(j), O_n(j) \in S_0).$$

Then, Procedure 2 provides asymptotic control of the family-wise error rate at level α , that is,

$$\limsup_{n \rightarrow \infty} \Pr(V_n > 0) \leq \alpha.$$

If condition (32) in Assumption AP1 holds with equality and Q_0 is continuous (so that $\min_{j \in S_0} \bar{Q}_{0j}(Z(j))$ is a continuous random variable for $Z \sim Q_0$), then asymptotic control is exact

$$\lim_{n \rightarrow \infty} \Pr(V_n > 0) = \alpha.$$

Proof of Theorem 5. The proof is analogous to that of Theorem 2, thus we only highlight the main steps where Assumptions AP1 and AP2 come into play. Compared to the previous proof, maxima of test statistics are replaced

by minima of unadjusted p -values and the direction of the cut-off rules are reversed. Again, the procedure can also be stated equivalently in terms of statistics $P_{0n}^*(j)$, as follows. Let

$$\begin{aligned} P_{0n}^*(1) &\equiv P_{0n}^\circ(1) \\ P_{0n}^*(j) &\equiv \begin{cases} P_{0n}^\circ(j), & \text{if } P_{0n}^*(j-1) < C_n(j-1) \\ 1, & \text{otherwise} \end{cases}, \quad j = 2, \dots, m, \end{aligned}$$

and reject $H_{0,O_n(j)}$ if $P_{0n}^*(j) < C_n(j)$, $j = 1, \dots, m$. As before, define Bernoulli random variables

$$B_n \equiv \mathbb{I}(\{O_n(1), \dots, O_n(m_1)\} = S_0^c, P_{0n}^*(1) < C_n(1), \dots, P_{0n}^*(m_1) < C_n(m_1))$$

and argue that, under asymptotic separation Assumption AP2, then $Pr(B_n = 1) \rightarrow 1$ as $n \rightarrow \infty$. Asymptotic null domination Assumption AP1 comes into play at the very last step of the proof to show that

$$\limsup_{n \rightarrow \infty} Pr\left(\min_{j \in S_0} P_{0n}(j) < c(S_0)\right) \leq \alpha.$$

□

2.2.3 Explicit proposal for the test statistics null distribution

A null distribution Q_0 for Procedure 2 can be constructed as described in Theorem 3, with a few additional requirements in order to meet Assumptions AP1 and AP2 of Theorems 4 and 5. Lemma 1 and 2 are concerned with providing sufficient conditions (in terms of continuity and monotonicity assumptions on the null distribution Q_0) so that Assumptions AP1 and AP2 are implied by their maxT counterparts, i.e., by Assumptions AT1 and AT2, respectively.

Lemma 1 [Asymptotic null domination] *The null distribution $Q_0 = Q_0(P)$ defined in Theorem 3, with the additional condition that the marginal survivor functions $\bar{Q}_{0j} = 1 - Q_{0j}$, $j = 1, \dots, m$, are continuous, satisfies Assumption AP1.*

Proof of Lemma 1. Define an intermediate random vector $(\tilde{Z}_n(j) : j \in S_0)$ as in the proof of Theorem 3, so that $T_n(j) \leq \tilde{Z}_n(j)$, for $j \in S_0$, and $(\tilde{Z}_n(j) :$

$j \in S_0$) and $(Z_n(j) : j \in S_0)$ have the same limit distribution Q_{0,S_0} . Then, by the Continuous Mapping Theorem,

$$\begin{aligned}
\limsup_{n \rightarrow \infty} Pr \left(\min_{j \in S_0} P_{0n}(j) < x \right) &= \limsup_{n \rightarrow \infty} Pr \left(\min_{j \in S_0} \bar{Q}_{0j}(T_n(j)) < x \right) \\
&\leq \limsup_{n \rightarrow \infty} Pr \left(\min_{j \in S_0} \bar{Q}_{0j}(\tilde{Z}_n(j)) < x \right) \\
&= Pr \left(\min_{j \in S_0} \bar{Q}_{0j}(Z(j)) < x \right) \\
&= Pr \left(\min_{j \in S_0} P_0(j) < x \right) \text{ for all } x,
\end{aligned}$$

where $Z \sim Q_0$. In particular, if (17) holds with equality, then (32) also holds with equality.

□

Lemma 2 [Asymptotic separation of true and false null hypotheses]

Suppose that the marginal survivor functions $\bar{Q}_{0j}$, $j = 1, \dots, m$, corresponding to the null distribution Q_0 satisfy the following: (i) $\bar{Q}_{0j}$ is continuous; (ii) $\bar{Q}_{0j}$ is strictly decreasing; and (iii) there exists an M_1 (possibly degenerate) such that $\lim_{\epsilon \downarrow 0} \bar{Q}_{0j}^{-1}(\epsilon) = M_1$ for each j . Then, asymptotic separation Assumption AT2 for the test statistics $T_n(j)$ implies Assumption AP2 for the unadjusted p -values $P_{0n}(j) = \bar{Q}_{0j}(T_n(j))$, $j = 1, \dots, m$.

Proof of Lemma 2. Condition (33) follows from (12) by noting that, for each $\epsilon > 0$, $M(\epsilon) \equiv \max_{j \in S_0^c} \bar{Q}_{0j}^{-1}(\epsilon) < M_1$, and

$$\begin{aligned}
Pr \left(\max_{j \in S_0^c} P_{0n}(j) \leq \epsilon \right) &= Pr \left(\bar{Q}_{0j}(T_n(j)) \leq \epsilon, \forall j \in S_0^c \right) \\
&= Pr \left(T_n(j) \geq \bar{Q}_{0j}^{-1}(\epsilon), \forall j \in S_0^c \right) \\
&\geq Pr \left(T_n(j) \geq M(\epsilon), \forall j \in S_0^c \right) \\
&= Pr \left(\min_{j \in S_0^c} T_n(j) \geq M(\epsilon) \right).
\end{aligned}$$

Similarly, condition (34) follows from (13) by noting that, for each $\epsilon > 0$ and $m(\epsilon) \equiv \min_{j \in S_0} \bar{Q}_{0j}^{-1}(\epsilon)$, then

$$\begin{aligned} Pr \left(\min_{j \in S_0} P_{0n}(j) \leq \epsilon \right) &= Pr \left(\bar{Q}_{0j}(T_n(j)) \leq \epsilon, \text{ for some } j \in S_0 \right) \\ &= Pr \left(T_n(j) \geq \bar{Q}_{0j}^{-1}(\epsilon), \text{ for some } j \in S_0 \right) \\ &\leq Pr \left(T_n(j) \geq m(\epsilon), \text{ for some } j \in S_0 \right) \\ &= Pr \left(\max_{j \in S_0} T_n(j) \geq m(\epsilon) \right). \end{aligned}$$

Also, $m(\epsilon) \uparrow M_1$ as $\epsilon \downarrow 0$, thus, by (13),

$$\lim_{\epsilon \downarrow 0} \lim_{n \rightarrow \infty} Pr \left(\min_{j \in S_0} P_{0n}(j) \leq \epsilon \right) \leq \lim_{\epsilon \downarrow 0} \lim_{n \rightarrow \infty} Pr \left(\max_{j \in S_0} T_n(j) \geq m(\epsilon) \right) = 0.$$

□

2.2.4 Adjusted p -values

Adjusted p -values are obtained similarly as for Procedure 1. Again, for simplicity (and as in Lemma 1 and 2), consider a null distribution Q_0 with continuous and strictly monotone marginal c.d.f.'s, Q_{0j} , and survivor functions, $\bar{Q}_{0j} = 1 - Q_{0j}$, $j = 1, \dots, m$.

Result 2 [Adjusted p -values for step-down minP Procedure 2] *The adjusted p -values for step-down minP Procedure 2, based on a null distribution Q_0 with continuous and strictly monotone marginal distributions, are given by*

$$\tilde{P}_{0n}(O_n(j)) = \max_{k=1, \dots, j} \left\{ Pr_{Q_0} \left(\min_{l \in \{O_n(k), \dots, O_n(m)\}} P_0(l) \leq P_{0n}(O_n(k)) \right) \right\}, \quad (36)$$

where $P_0(j) = \bar{Q}_{0j}(Z(j))$, $Z = (Z(j) : j = 1, \dots, m) \sim Q_0$, and $H_{0, O_n(j)}$ is the null hypothesis corresponding to the j th most significant unadjusted p -value $P_{0n}^\circ(j) = P_{0n}(O_n(j)) = \bar{Q}_{0, O_n(j)}(T_n(O_n(j)))$, that is, the indices $O_n(j)$ are defined such that $P_{0n}(O_n(1)) \leq \dots \leq P_{0n}(O_n(m))$. Step-down Procedure 2 for controlling the FWER at level α can then be stated equivalently as

$$S(T_n, Q_0, \alpha) = \{O_n(j) : \tilde{P}_{0n}(O_n(j)) \leq \alpha, j = 1, \dots, m\}.$$

Proof of Result 2. As in Procedure 2, let $F_{A,Q_0}(z) \equiv \Pr_{Q_0}(\min_{j \in A} P_0(j) \leq z)$ denote the c.d.f. of $\min_{j \in A} P_0(j)$ for $Z \sim Q_0$, $\bar{O}_n(j) \equiv \{O_n(j), \dots, O_n(m)\}$, and $C_n(j) = F_{\bar{O}_n(j), Q_0}^{-1}(\alpha)$. Then,

$$\begin{aligned} \tilde{P}_{0n}(O_n(j)) &= \inf \left\{ \alpha \in [0, 1] : \sum_{k=1}^j \mathbf{I}(P_{0n}(O_n(k)) < C_n(k)) = j \right\} \\ &= \inf \{ \alpha \in [0, 1] : P_{0n}(O_n(k)) < C_n(k), \forall k = 1, \dots, j \} \\ &= \max_{k=1, \dots, j} \inf \{ \alpha \in [0, 1] : P_{0n}(O_n(k)) < C_n(k) \} \\ &= \max_{k=1, \dots, j} \inf \left\{ \alpha \in [0, 1] : P_{0n}(O_n(k)) < F_{\bar{O}_n(k), Q_0}^{-1}(\alpha) \right\} \\ &= \max_{k=1, \dots, j} \inf \left\{ \alpha \in [0, 1] : F_{\bar{O}_n(k), Q_0}(P_{0n}(O_n(k))) \leq \alpha \right\} \\ &= \max_{k=1, \dots, j} F_{\bar{O}_n(k), Q_0}(P_{0n}(O_n(k))) \\ &= \max_{k=1, \dots, j} \left\{ \Pr_{Q_0} \left(\min_{l \in \{O_n(k), \dots, O_n(m)\}} P_0(l) \leq P_{0n}(O_n(k)) \right) \right\}. \end{aligned}$$

□

These adjusted p -values correspond to those given in equation (2.10), p. 66, of Westfall and Young (1993), again with an important distinction in the choice of the null distribution Q_0 . Consider the special case where the random m -vector $Z \sim Q_0$ has *independent* components $Z(j)$, with continuous marginal distributions Q_{0j} , $j = 1, \dots, m$. Then, the unadjusted p -values $P_0(j) = \bar{Q}_{0j}(Z(j))$ are independent $U(0, 1)$ random variables and the minima $\min_{j \in A} P_0(j)$ have $Beta(1, |A|)$ distributions. The adjusted p -values for Procedure 2 then reduce to the *step-down Šidák adjusted p -values* (Dudoit et al., 2003a)

$$\tilde{P}_{0n}(O_n(j)) = \max_{k=1, \dots, j} \left\{ 1 - (1 - P_{0n}(O_n(k)))^{(m-k+1)} \right\}. \quad (37)$$

Thus, in this independence situation, the step-down minP procedure is very simple and is based only on the marginal null distributions, Q_{0j} . In general, however, the test statistics are not independent and Procedure 2, based on a null distribution Q_0 constructed as in Theorem 3, takes into account the *joint* distribution of the test statistics when computing quantiles $c(A, Q_0, \alpha)$ and the resulting adjusted p -values $\tilde{P}_{0n}(O_n(j))$.

3 Bootstrap-based step-down procedures for control of the family-wise error rate

In practice, since the data generating distribution P is unknown, then so is the null distribution $Q_0 = Q_0(P)$ defined in Theorem 3. Estimation of Q_0 is then needed, especially to deal with the unknown dependence structure among the test statistics. In this section, we consider analogues of Procedures 1 and 2, based on a consistent estimator Q_{0n} of a null distribution Q_0 , such as that defined in Theorem 3. In such multiple testing procedures, the estimator Q_{0n} is used in place of Q_0 , to estimate the cut-offs for the test statistics and the resulting adjusted p -values. A more detailed discussion of different estimation methods, including test statistics specific approaches, is provided in Dudoit et al. (2003b).

3.1 Asymptotic control for consistent estimator of the null distribution

Theorem 6 [Consistency of step-down maxT cut-offs in Procedure 1] *Let Q_0 be a specified m -variate null distribution and let Q_{0n} converge weakly to Q_0 . For an arbitrary m -variate distribution Q , random m -vector $Z = (Z(j) : j = 1, \dots, m) \sim Q$, and level $\alpha \in (0, 1)$, define $(1 - \alpha)$ -quantiles, $c(A, Q, \alpha) \in \mathbb{R}$, for the distributions of maxima, $\max_{j \in A} Z(j)$, of random variables $Z(j)$ over subsets $A \subseteq \{1, \dots, m\}$,*

$$c(A, Q, \alpha) \equiv F_{A,Q}^{-1}(1 - \alpha) = \inf \{z : F_{A,Q}(z) \geq 1 - \alpha\},$$

where $F_{A,Q}(z) \equiv \Pr_Q(\max_{j \in A} Z(j) \leq z)$ denotes the c.d.f. of $\max_{j \in A} Z(j)$ for $Z \sim Q$. In particular, for the null distribution Q_0 , assume that for each subset $A \subseteq \{1, \dots, m\}$, F_{A,Q_0} is continuous and has Lebesgue density f_{A,Q_0} with interval support, that is, $\{z : f_{A,Q_0}(z) > 0\} = (a_A, b_A)$, where a_A and b_A are allowed to equal $-\infty$ and ∞ , respectively. Then, one has the following consistency result for the step-down maxT cut-offs

$$\lim_{n \rightarrow \infty} c(A, Q_{0n}, \alpha) = c(A, Q_0, \alpha), \quad \forall A \subseteq \{1, \dots, m\}.$$

Proof of Theorem 6. Consider random m -vectors $Z_n \sim Q_{0n}$ and $Z \sim Q_0$. We have that $(Z_n(j) : j \in A)$ converges weakly to $(Z(j) : j \in A)$, $\forall A \subseteq \{1, \dots, m\}$. In particular, by the Continuous Mapping Theorem, this

implies that $\max_{j \in A} Z_n(j)$ converges weakly to $\max_{j \in A} Z(j)$, so that $F_{A, Q_{0n}}$ converges pointwise to F_{A, Q_0} at each continuity point of F_{A, Q_0} . Since pointwise convergence of monotone functions to a continuous monotone function implies uniform convergence, this proves that $F_{A, Q_{0n}}$ converges uniformly to F_{A, Q_0} . By continuity of the quantile mapping $F \rightarrow F^{-1}(1 - \alpha)$, with respect to the supremum norm convergence at F_{A, Q_0} with $f_{A, Q_0}(F_{A, Q_0}^{-1}(1 - \alpha)) > 0$ (where we use that, by assumption, $F_{A, Q_0}^{-1}(1 - \alpha) \in (a_A, b_A)$), this proves that $F_{A, Q_{0n}}^{-1}(1 - \alpha) = c(A, Q_{0n}, \alpha)$ converges to $F_{A, Q_0}^{-1}(1 - \alpha) = c(A, Q_0, \alpha)$, as $n \rightarrow \infty$, $\forall A \subseteq \{1, \dots, m\}$.

□

Note that the above proof corresponds to the proof of Theorem 4, for consistency of the single-step common cut-offs in Procedure 2 of Dudoit et al. (2003b), with the following modifications: (i) θ is the FWER-specific mapping, $\theta(F) = 1 - F(0)$; and (ii) the number of rejected hypotheses R is computed over subsets $A \subseteq \{1, \dots, m\}$, rather than over the entire set $\{1, \dots, m\}$, that is, one considers A -specific numbers of rejections, $R_A((c, \dots, c) \mid Q) \equiv \sum_{j \in A} \mathbf{I}(Z(j) > c)$, for $Z \sim Q$. One can then define A -specific functions, $c \rightarrow G_{A, Q}(c) \equiv \theta(F_{R_A((c, \dots, c) \mid Q)})$, and note that $G_{A, Q}(c) = 1 - F_{A, Q}(c)$, so that assumptions regarding G_{Q_0} , in single-step Theorem 4, translate into assumptions on the c.d.f. F_{A, Q_0} of $\max_{j \in A} Z(j)$, in step-down Theorem 6, above.

Consistency of the step-down minP cut-offs for Procedure 2 follows from Theorem 3, on consistency of the single-step common-quantile cut-offs in Procedure 1 of Dudoit et al. (2003b), with modifications (i) and (ii), above. A general consistency result for A -specific common quantiles is stated below for arbitrary Type I error rate mappings $\theta(\cdot)$. The proof is identical to that of Theorem 3 in Dudoit et al. (2003b), but with A -specific numbers of rejections R_A , and is therefore omitted here.

Theorem 7 [Consistency of A -specific common quantiles] *Let Q_0 be a specified m -variate null distribution and let Q_{0n} converge weakly to Q_0 . Assume that Q_0 is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^m$, with uniformly bounded density, and that each marginal distribution Q_{0j} has continuous Lebesgue density f_{0j} with interval support, that is, $\{z : f_{0j}(z) > 0\} = (a_j, b_j)$, where a_j and b_j are allowed to equal $-\infty$ and ∞ ,*

respectively. For an arbitrary m -variate distribution Q and constant $\delta \in [0, 1]$, define δ -quantiles for the marginal distributions Q_j by

$$d_j(Q, \delta) \equiv Q_j^{-1}(\delta) = \inf\{z : Q_j(z) \geq \delta\}, \quad j = 1, \dots, m, \quad (38)$$

and let $d_A(Q, \delta) = (d_j(Q, \delta) : j \in A)$ denote the corresponding quantile vectors for subsets $A \subseteq \{1, \dots, m\}$. Define functions

$$\delta \rightarrow G_{A,Q}(\delta) \equiv \theta \left(F_{R_A(d_A(Q, \delta) | Q)} \right), \quad (39)$$

where

$$R_A(d_A(Q, \delta) | Q) \equiv \sum_{j \in A} I(Z(j) > d_j(Q, \delta))$$

is the number of A -specific rejected hypotheses for $Z \sim Q$. For a fixed level $\alpha \in (0, 1)$ and any subset $A \subseteq \{1, \dots, m\}$, define

$$\delta(A, Q) \equiv G_{A,Q}^{-1}(\alpha) = \inf \left\{ \delta : \theta \left(F_{R_A(d_A(Q, \delta) | Q)} \right) \leq \alpha \right\}. \quad (40)$$

In particular, for the null distribution Q_0 , assume that $\delta(A, Q_0) \in (0, 1)$ and that the function $G_{A,Q_0}(\delta)$ is continuous and has a positive derivative at $\delta(A, Q_0) = G_{A,Q_0}^{-1}(\alpha)$. Then, one has the following consistency result for the A -specific common quantiles. For each $A \subseteq \{1, \dots, m\}$, as $n \rightarrow \infty$,

$$\begin{aligned} \delta(A, Q_{0n}) - \delta(A, Q_0) &\rightarrow 0 \quad \text{and} \\ d_j(Q_{0n}, \delta(A, Q_{0n})) - d_j(Q_0, \delta(A, Q_0)) &\rightarrow 0, \quad \forall j = 1, \dots, m. \end{aligned}$$

Theorem 8, below, shows that consistency of the step-down minP cut-offs follows from Theorem 7, by noting that these cut-offs are equal to the δ 's of Theorem 7, that is, $c(A, Q, \alpha) = \delta(A, Q)$.

Theorem 8 [Consistency of step-down minP cut-offs in Procedure 2] Let Q_0 be an m -variate null distribution as in Theorem 7 and let Q_{0n} converge weakly to Q_0 . For an arbitrary m -variate distribution Q , random m -vector $Z = (Z(j) : j = 1, \dots, m) \sim Q$, and level $\alpha \in (0, 1)$, define α -quantiles, $c(A, Q, \alpha) \in [0, 1]$, for the distributions of minima, $\min_{j \in A} \bar{Q}_j(Z(j))$, of unadjusted p -values $P(j) = \bar{Q}_j(Z(j))$ over subsets $A \subseteq \{1, \dots, m\}$,

$$c(A, Q, \alpha) \equiv F_{A,Q}^{-1}(\alpha) = \inf \{z : F_{A,Q}(z) \geq \alpha\}, \quad (41)$$

where Q_j and $\bar{Q}_j = 1 - Q_j$, $j = 1, \dots, m$, denote, respectively, the marginal c.d.f.'s and survivor functions corresponding to Q , and $F_{A,Q}(z) \equiv \Pr_Q \left(\min_{j \in A} \bar{Q}_j(Z(j)) \leq z \right)$

denotes the c.d.f. of $\min_{j \in A} \bar{Q}_j(Z(j))$ for $Z \sim Q$. For an m -variate distribution Q , with continuous and strictly increasing marginal c.d.f.'s Q_j , $G_{A,Q}(\delta) = F_{A,Q}(1 - \delta)$, so that the step-down minP cut-offs are equal to the $\delta(A, Q)$'s of Theorem 7, that is,

$$c(A, Q, \alpha) = \delta(A, Q), \quad \forall A \subseteq \{1, \dots, m\}.$$

Consequently, by Theorem 7, we have the following consistency result for the step-down minP cut-offs

$$\lim_{n \rightarrow \infty} c(A, Q_{0n}, \alpha) = c(A, Q_0, \alpha), \quad \forall A \subseteq \{1, \dots, m\}.$$

Proof of Theorem 8. For an m -variate distribution Q , with continuous and strictly increasing marginal c.d.f.'s Q_j ,

$$\begin{aligned} G_{A,Q}(\delta) &= 1 - F_{R_A(d_A(Q,\delta)|Q)}(0) \\ &= Pr_Q \left(\sum_{j \in A} I(Z(j) > d_j(Q, \delta)) > 0 \right) \\ &= Pr_Q (\exists j \in A, Z(j) > d_j(Q, \delta)) \\ &= Pr_Q (\exists j \in A, \bar{Q}_j(Z(j)) \leq \bar{Q}_j(d_j(Q, \delta))) \\ &= Pr_Q (\exists j \in A, \bar{Q}_j(Z(j)) \leq 1 - \delta) \\ &= Pr_Q \left(\min_{j \in A} \bar{Q}_j(Z(j)) \leq 1 - \delta \right) \\ &= F_{A,Q}(1 - \delta), \end{aligned}$$

where $F_{A,Q}$ denotes the c.d.f. of $\min_{j \in A} \bar{Q}_j(Z(j))$ for $Z \sim Q$. Hence,

$$\begin{aligned} c(A, Q, \alpha) &= \inf \{z \in [0, 1] : F_{A,Q}(z) \geq \alpha\} \\ &= \sup \{z \in [0, 1] : F_{A,Q}(z) \leq \alpha\} \\ &= \sup \{z \in [0, 1] : G_{A,Q}(1 - z) \leq \alpha\} \\ &= \inf \{\delta \in [0, 1] : G_{A,Q}(\delta) \leq \alpha\} \\ &= \delta(A, Q). \end{aligned}$$

Convergence of the estimated step-down minP cut-offs $c(A, Q_{0n}, \alpha)$ to $c(A, Q_0, \alpha)$ is then a direct consequence of the convergence of $\delta(A, Q_{0n})$ to $\delta(A, Q_0)$, as established in Theorem 7.

□

For a broad class of testing problems, the null distribution $Q_0 = Q_0(P)$, as constructed in Theorem 3, has continuous and strictly monotone marginal distributions. For example, for the test of single-parameter null hypotheses using t -statistics, $Q_0(P)$ is an m -variate Gaussian distribution with mean vector zero (Section 5.1 in Dudoit et al. (2003b)). In such cases, consistent estimators Q_{0n} can also be defined in terms of Gaussian distributions, with a suitable estimator of the covariance matrix. The assumptions of Theorem 8 are therefore satisfied by both Q_0 and Q_{0n} . In the case when Q_{0n} is not continuous (e.g., obtained from general bootstrap Procedure 3, below), but converges in distribution to a continuous Q_0 , Theorem 8 strongly suggests asymptotic equality of $c(A, Q_{0n}, \alpha)$ and $\delta(A, Q_{0n})$.

Having established consistency of the cut-offs for step-down Procedures 1 and 2, based on a consistent estimator Q_{0n} of the null distribution Q_0 , Corollary 1 from Dudoit et al. (2003b) can be applied to prove consistency of the resulting Type I error rates.

3.2 Bootstrap estimation of the null distribution

The null distribution $Q_0 = Q_0(P)$ of Theorem 3 can be estimated with the non-parametric or model-based bootstrap. Let P_n^* denote an estimator of the true data generating distribution P . For the *non-parametric bootstrap*, P_n^* is simply the empirical distribution P_n , that is, samples of size n are drawn at random with replacement from the observed $X_1, \dots, X_n$. For the *model-based bootstrap*, P_n^* is based on a model $\mathcal{M}$ for the data generating distribution P , such as the family of m -variate Gaussian distributions.

A bootstrap sample consists of n i.i.d. realizations, $X_1^\#, \dots, X_n^\#$, of a random variable $X^\# \sim P_n^*$. Denote the m -vector of test statistics computed from such a bootstrap sample by $T_n^\# = (T_n^\#(j) : j = 1, \dots, m)$. The null distribution Q_0 proposed in Theorem 3 can be estimated by the distribution of the null-value shifted and scaled bootstrap statistics

$$Z_n^\#(j) \equiv \sqrt{\min\left(1, \frac{\tau_0(j)}{\text{Var}_{P_n^*}[T_n^\#(j)]}\right)} \left(T_n^\#(j) + \lambda_0(j) - E_{P_n^*}[T_n^\#(j)]\right). \quad (42)$$

In practice, one can only approximate the distribution of $Z_n^\# = (Z_n^\#(j) : j =$

$1, \dots, m)$ by an empirical distribution over B bootstrap samples drawn from P_n^* , as described next in Procedure 3.

Procedure 3. Bootstrap-based step-down procedures for FWER control.

1. Generate B bootstrap samples, $(X_1^b, \dots, X_n^b)$, $b = 1, \dots, B$. For the b th sample, the X_i^b , $i = 1, \dots, n$, are n i.i.d. realizations of a random variable $X^\# \sim P_n^*$.
2. For each bootstrap sample, compute an m -vector of test statistics, $T_n^b = (T_n^b(j) : j = 1, \dots, m)$, which can be arranged in an $m \times B$ matrix, $\mathbf{T} = (T_n^b(j))$, with rows corresponding to the m hypotheses and columns to the B bootstrap samples.
3. Compute row means and variances of the matrix $\mathbf{T}$, to yield estimates of $E[T_n(j)]$ and $Var[T_n(j)]$, $j = 1, \dots, m$.
4. Obtain an $m \times B$ matrix, $\mathbf{Z} = (Z_n^b(j))$, of null-value shifted and scaled bootstrap statistics $Z_n^b(j)$, as in Theorem 3, by row-shifting and scaling the matrix $\mathbf{T}$ using the bootstrap estimates of $E[T_n(j)]$ and $Var[T_n(j)]$ and the user-supplied null-values $\lambda_0(j)$ and $\tau_0(j)$.
5. The bootstrap estimate Q_{0n} of the null distribution Q_0 from Theorem 3 is the empirical distribution of the columns Z_n^b of matrix $\mathbf{Z}$.
6. For **step-down maxT Procedure 1**, the estimated quantiles $c(A, Q_{0n}, \alpha)$ are simply the $(1 - \alpha)$ -quantiles of $\max_{j \in A} Z_n^b(j)$ over the B bootstrap samples, that is,

$$c(A, Q_{0n}, \alpha) \equiv \inf \left\{ z : \frac{1}{B} \sum_{b=1}^B \mathbf{I} \left(\max_{j \in A} Z_n^b(j) \leq z \right) \geq 1 - \alpha \right\}.$$

7. For **step-down minP Procedure 2**, one must first estimate unadjusted p -values using Q_{0n} , before considering the distribution of their successive minima. Estimated unadjusted p -values are obtained by row-ranking the matrix $\mathbf{Z}$ and are given by

$$\bar{Q}_{0n,j}(T_n(j)) \equiv \frac{1}{B} \sum_{b=1}^B \mathbf{I}(Z_n^b(j) > T_n(j)).$$

The reader is referred to Ge et al. (2003) for a fast algorithm for computing resampling-based (bootstrap or permutation) adjusted p -values for step-down minP procedures.

References

- 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/.
- Y. Ge, S. Dudoit, and T. P. Speed. Resampling-based multiple testing for microarray data analysis. *TEST*, 12(1), 2003.
- 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).
- M. J. van der Laan, S. Dudoit, and K. S. Pollard. Multiple testing. Part III. Procedures for control of the generalized family-wise error rate and proportion of false positives. Technical Report 140, Division of Biostatistics, UC Berkeley, 2003. URL www.bepress.com/ucbbiostat/paper140/.
- P. H. Westfall and S. S. Young. *Resampling-based multiple testing: Examples and methods for p-value adjustment*. John Wiley & Sons, 1993.