ELSEVIER

Journal of Statistical Planning and Inference 84 (2000) 139–158

journal of statistical planning and inference

www.elsevier.com/locate/jspi

Stepwise normal theory multiple test procedures controlling the false discovery rate

James F. Troendle[a,b,*]

[a] *Biometry and Mathematical Statistics Branch, Division of Epidemiology, Statistics, and Prevention Research, National Institute of Child Health and Human Development, USA*

[b] *NIH, Bld. 6100 Rm. 7B13, Bethesda, MD 20892, USA*

Received 1 September 1998; received in revised form 20 May 1999; accepted 28 June 1999

Abstract

The false discovery rate (FDR), or the expected proportion of falsely rejected null hypotheses to rejected null hypotheses, has recently been proposed as an error rate that multiple testing procedures should in certain circumstances control. So far, only a step-up procedure for independent test statistics has been created explicitly to control the FDR (Benjamini and Hochberg, 1995). In this paper, step-down and step-up procedures are described which asymptotically (as $N \to \infty$) control the FDR when the test statistics are the t statistics from consistent multivariate normal estimators of the tested parameters. Determination of the necessary critical constants for the normal theory procedures is achieved using numerical integration when the correlations are equal, or through simulation using the multivariate t distribution when the correlations are arbitrary. The critical constants of the normal theory procedures are compared to those of the Benjamini and Hochberg procedure under the normal assumption, and a large potential power increase is found. Simulation strongly supports the use of critical constants, obtained by an asymptotic argument, in small samples for as many as 30 tests. Adjusted FDR values can be found to quantify the evidence against a given hypothesis.

MSC: 62F03

Keywords: False discovery rate; Multiple testing; Step-down; Step-up; Power

1. Introduction

Benjamini and Hochberg (1995) introduce a new error rate for multiple testing procedures. The false discovery rate (FDR) is defined as the expectation of the proportion of errors among the rejected hypotheses, with the understanding that if no null hypotheses are rejected the proportion is zero. Traditionally, control of the familywise

* Corresponding author. NIH, Bld. 6100 Rm. 7B13, Bethesda, MD 20892, USA.
E-mail address: jt3t@nih.gov (J.F. Troendle)

0378-3758/00/$ - see front matter
PII: S0378-3758(99)00145-7

error rate (FWER) is adopted as the criterion for a multiple comparison procedure. The FWER for a procedure is defined as the probability of committing any type I error among all comparisons made by the procedure. The FDR is less stringent than the FWER and so any procedure controlling the FWER also controls the FDR. This allows FDR controlling procedures to have greater power than FWER controlling procedures, which may lead more researchers to actually use multiplicity adjustment.

Use of the FDR should be tempered considerably by the observation that adding several false null hypotheses that are sure to be rejected will decrease the FDR. Therefore, a procedure that controls the FDR may reject hypotheses with arbitrarily large unadjusted p-values in some extreme cases. This could lead a researcher to add variables for which there was strong evidence against the corresponding null hypothesis, thereby adding to the chance of rejecting other hypotheses. This is certainly a large concern. But if the variables to be studied are not specified before data collection, all kinds of problems are possible. In FWER control, a dishonest researcher can remove from the testing procedure the variables for which there is not strong evidence against their corresponding null hypotheses. This is effectively sidestepping the multiple comparison issue, and should be strongly discouraged. Nevertheless, researchers do use the FWER. While the FWER is the most appropriate error rate to control in confirmatory analyses, there are exploratory analyses where less control is desired in exchange for more power. Also, notice that the gain in using the FDR compared to the FWER is increasing when the number of false null hypotheses is increasing. Consequently, it seems that the FDR is well suited for testing situations in which there are many different treatments or outcomes for which there is little or no prior knowledge, and the objective is to identify a smaller set of them to study further. In such screening situations, no adjustment may result in too many treatments being considered at the costly confirmatory stage. Thus, the FDR represents a compromise between too much or too little adjustment in screening applications. Examples include drug screening tests to identify potential drugs for development, and testing multiple factors in an experimental design. Another case is when there are many different measures of outcome, where one wants to know which are significantly increased or decreased by a particular treatment. In order to make an overall decision about the treatment, it is not necessary that every individual decision be correct. This is a case in which the FDR may be used.

Benjamini and Hochberg (1995) give a step-up (BHSU) multiple testing procedure which controls the FDR if the test statistics used are independent. Suppose that one wishes to test the null hypotheses $H_1, H_2, \ldots, H_k$ with the normalized (so that the order of the test statistics is the reverse of the order of the p-values) test statistics $T_1, T_2, \ldots, T_k$ and corresponding p-values $P_1, P_2, \ldots, P_k$. Order the test statistics $T_{(1)} \leqslant T_{(2)} \leqslant \cdots \leqslant T_{(k)}$ and denote by $P_{(i)}$ and $H_{(i)}$ the p-value and null hypothesis corresponding to $T_{(i)}$. Benjamini and Hochberg's test is a step-up procedure, meaning that one starts by testing $H_{(1)}$, the hypothesis corresponding to the smallest test statistic (largest p-value), and sequentially accept null hypotheses until the first time that $T_{(i)} \geqslant c_i$, or equivalently $P_{(i)} \leqslant \alpha_i$, at which point $H_{(i)}, H_{(i+1)}, \ldots, H_{(k)}$ are rejected. The

critical cutoff values $\{\alpha_i\}$ for the BHSU test are given by

$$\alpha_i = \frac{(k-i+1)\alpha}{k}.$$

The resulting procedure controls the FDR at α if $T_1, T_2, \ldots, T_k$ are independent, or if the statistics have positive regression dependency on each of the statistics corresponding to true null hypotheses (Benjamini and Yekutiely, 1998).

The BHSU procedure is simple to use and provides FDR control regardless of the distribution type (as long as the test statistics have proper level for the individual hypothesis tests). However, the correlation structure of the data should be incorporated to increase the power of the procedure. Yekutiely and Benjamini (1999) use resampling to incorporate the dependence in a FDR controlling procedure. Similarly, the procedures given here use the correlation of the test statistics to produce more exact adjustments, leading to higher power. Step-down and step-up procedures are obtained under the assumption of normally distributed parameter estimators. A step-down procedure starts by testing $H_{(k)}$, the hypothesis corresponding to the largest test statistic, and sequentially rejecting null hypotheses until the first time that $T_{(i)} < c_i$ at which point $H_{(1)}, H_{(2)}, \ldots, H_{(i)}$ are accepted. The development of step-down and step-up tests follows that of Dunnett and Tamhane (1992,1995), who considered the normal theory FWER control case. Initially, assume that the test statistics have t distributions with equal variances and that the known correlation is the same for all pairs. The analytic solution for the critical constants $\{c_i\}$ is complicated by unequal correlation, but can be approximated by simulation.

In Section 2, we shall describe the assumptions and notation needed to define the procedures. The one-sided step-down procedure is introduced in Section 3, and the step-up procedure follows in Section 4. A comparison of the critical constants for the FDR controlling procedures is made in Section 5. The validity of the new procedures is established in two steps. First, the procedures are constructed so as to asymptotically control the FDR. Second, the small sample FDR is observed through simulation to be controlled for a wide range of correlation. An extension to the unequal correlation case is made in Section 6. The problem of finding adjusted p-values, or in this case adjusted FDR-values, for the procedures is addressed in Section 7. Modifications necessary for tests against two-sided alternatives are given in Section 8. The discussion follows in Section 9.

2. Preliminaries

Consider testing the set of hypotheses H_i: $\theta_i \leqslant 0$ versus A_i: $\theta_i > 0$ for $i = 1, 2, \ldots, k$. The two-sided alternative case will be considered later in Section 8. Assume that unbiased estimators $\hat{\theta}_1, \ldots, \hat{\theta}_k$ are available, each based on a sample of size N, with a multivariate normal distribution, $\mathrm{var}(\hat{\theta}_i) = \tau^2\sigma^2$ and $\mathrm{corr}(\hat{\theta}_i, \hat{\theta}_j) = \rho$, where τ^2 and ρ are known constants and σ^2 is an unknown error variance. Let S^2 be an unbiased estimator of σ^2 having v degrees of freedom (df) such that vS^2/σ^2 has a χ^2_v distribution

independent of the $\hat{\theta}_i$. Examples of such designs are given in Dunnett and Tamhane (1992). Note that in addition to the linear model examples given there, the problem is applicable to the two group multiple outcome example given in Section 1 in which a t test is used to compare each outcome with a pooled estimate of the common variance.

Let $t_i = \hat{\theta}_i/(s\tau)$, where s is the observed value of S. Then under H_i for $i = 1,\ldots,k$, $t_1, t_2, \ldots, t_k$ are observations from k-variate central t statistics, $T_1, T_2, \ldots, T_k$, with v df and common correlation ρ. Without loss of generality, assume that m $(0 \leqslant m \leqslant k)$ null hypotheses are true. Further suppose that the hypotheses have been relabeled if necessary so that $T_1, \ldots, T_m$ correspond to the true null hypotheses. If the statistics $T_1, T_2, \ldots, T_k$ are consistent for the tests of $H_1, H_2, \ldots, H_k$

$$T_i \text{ consistent test statistic for the test of } H_i \text{ versus } A_i, \quad i = 1,\ldots,k, \tag{1}$$

then asymptotically $(N \to \infty)$ we will have the following ordering:

$$\begin{aligned} &T_{(1)} \leqslant T_{(2)} \leqslant \cdots \leqslant T_{(m)} \leqslant T_{(m+1)} \leqslant \cdots \leqslant T_{(k)} \quad \text{corresponding to} \\ &\quad H_{(1)}, H_{(2)}, \ldots, H_{(m)}, H_{(m+1)}, \ldots, H_{(k)}, \end{aligned} \tag{2}$$

where $H_{(1)}, H_{(2)}, \ldots, H_{(m)}$ are true null hypotheses and $H_{(m+1)}, \ldots, H_{(k)}$ are false null hypotheses. Denote $P_{r:s}(\cdot)$, $r \leqslant s$, to be the probability under any parameter configuration with $H_{(1)}, \ldots, H_{(r)}$ true and $H_{(r+1)}, \ldots, H_{(s)}$ false.

3. Step-down procedure

Step-down procedures have been described in general in Section 1. The normal theory step-down FDR controlling procedure (NSD) will be determined upon specification of the critical constants $c_1, c_2, \ldots, c_k$. This section will describe how the critical constants are computed for one-sided tests, and verify the asymptotic FDR control. Let R be the number of null hypothesis rejected by the procedure, and V be the number of true null hypotheses rejected $(0 \leqslant V \leqslant R \leqslant k)$. The FDR is then

$$\begin{aligned} \text{FDR} &= E\left(\frac{V}{R}\right) = \sum_{r=1}^{k} \frac{1}{r} E(V|R=r) P_{m:k}(R=r) \\ &= \sum_{r=1}^{k} \frac{1}{r} E(V|R=r) \\ &\quad \times P_{m:k}(T_{(k)} \geqslant c_k, T_{(k-1)} \geqslant c_{k-1}, \ldots, T_{(k-r+1)} \geqslant c_{k-r+1}, T_{(k-r)} < c_{k-r}), \end{aligned} \tag{3}$$

where the last condition $T_{(k-r)} < c_{k-r}$ is imposed only when $r < k$. Expression (3) can be seen to be asymptotically equivalent (as $N \to \infty$) to

$$\sum_{r=k-m+1}^{k} \frac{(r-k+m)}{r} P_{m:m}(T_{(m)} \geqslant c_m, \ldots, T_{(k-r+1)} \geqslant c_{k-r+1}, T_{(k-r)} < c_{k-r}). \tag{4}$$

Actually $c_i = c_i(v)$, yet we leave off this additional notation throughout the paper. Note that $v \to \infty$ as $N \to \infty$. The c_i in Eq. (4) could therefore be taken as $c_i(\infty)$ if we desire, but we do not because we want a procedure with small sample properties

as exact as possible. The asymptotic ordering is needed to simplify the problem of selecting the constants.

If $c_1, c_2, \ldots, c_k$ are chosen so that (4) is contained below α for any $m \in \{1, 2, \ldots, k\}$, then the FDR will asymptotically be held below α. The statistics $T_1, \ldots, T_m$ come from the central multivariate t-distribution (Johnson and Kotz, 1976) by assumption. Therefore, as in Dunnett and Tamhane (1992), we can represent the variables as

$$T_i = \frac{\sqrt{1-\rho}Z_i - \sqrt{\rho}Z_0}{U}, \quad i = 1, \ldots, k,$$

where Z_i, $i = 1, \ldots, k$, are independent $N(0,1)$ variables and U is an independent $\sqrt{\chi^2_v/v}$ variable. Expression (4) becomes

$$\sum_{r=k-m+1}^{k} \frac{r-k+m}{r} \int_0^\infty \int_0^\infty P(Z_{(m)} \geqslant d_m, Z_{(m-1)} \geqslant d_{m-1}, \ldots, Z_{(k-r+1)} \geqslant d_{k-r+1}, Z_{(k-r)} < d_{k-r}) \phi(z_0)\, dz_0\, f_v(u)\, du, \tag{5}$$

where

$$d_i = \frac{c_i u + \sqrt{\rho} z_0}{\sqrt{1-\rho}}, \quad 1 \leqslant i \leqslant k,$$

$\phi(\cdot)$ is the standard normal density function, and $f_v(\cdot)$ is the density function of U. By symmetry, the condition in the probability term of (5) that $Z_{(k-r)} < d_{k-r}$ is equivalent to choosing $k-r$ Z's from $Z_1, \ldots, Z_m$ which must be less than d_{k-r}. Thus, we get for the probability term in (5).

$$\binom{m}{k-r} P(Z_1 < d_{k-r}, \ldots, Z_{k-r} < d_{k-r}) P(Z_{(m-k+r)} \geqslant d_m, \ldots, Z_{(1)} \geqslant d_{k-r+1})$$
$$= \binom{m}{k-r} [\Phi(d_{k-r})]^{k-r} F_{m-k+r}(-d_m, \ldots, -d_{k-r+1}), \tag{6}$$

where we have used the symmetry of the normal distribution, and where $F_m(x_1, \ldots, x_m) = P(Z_{(1)} < x_1, \ldots, Z_{(m)} < x_m)$. By conditioning on which order statistic is the first to exceed the corresponding $-d_i$, we get

$$F_{m-k+r}(-d_m, \ldots, -d_{k-r+1}) = 1 - \sum_{j=0}^{m-k+r-1} \binom{m-k+r}{j} F_j(-d_m, \ldots, -d_{m-j+1}) [\Phi(d_{m-j})]^{m-k+r-j}, \tag{7}$$

where $F_0 = 1$. Note that $F_{m-k+r}(\cdot, \ldots, \cdot)$ can be evaluated recursively. Expression (5) can then be determined by using numerical integration. We have adopted the same numerical strategy used by Dunnett and Tamhane (see Dunnett and Tamhane, 1990).

The first critical constant is determined by setting (4) with $m = 1$ equal to α,

$$\alpha = \frac{1}{k} P_{1:1}(T_{(1)} \geqslant c_1),$$

which has solution $c_1 = t_v(k\alpha)$, the upper $k\alpha$ point of the t distribution with v df. If $k\alpha$ exceeds $1/2$, then c_1 is taken to be zero by convention so as not to allow the

possibility of rejecting based on a negative t value. This is an arbitrary decision, but one that seems to be the most natural option available. This decision relates to the point discussed in Section 1, that in extreme cases an FDR procedure could reject hypotheses with arbitrarily large unadjusted p-values. Given c_1, we obtain c_2 by setting (5) with $m=2$ equal to α,

$$\alpha = \frac{1}{k-1}\int_0^\infty \int_{-\infty}^\infty P(Z_{(2)} \geqslant d_2, Z_{(1)} < d_1)\phi(z_0)\,\mathrm{d}z_0\, f_v(u)\,\mathrm{d}u + \frac{2}{k}\int_0^\infty \int_{-\infty}^\infty P(Z_{(2)} \geqslant d_2, Z_{(1)} \geqslant d_1)\phi(z_0)\,\mathrm{d}z_0\, f_v(u)\,\mathrm{d}u$$

and solving for $c_2 \in [c_1, \infty)$. The integrals are computed via numerical integration. A unique solution will exist as long as $k \leqslant 1/\alpha$. If $k > 1/\alpha$, then c_2 is taken to equal c_1, and the FDR is conservative (controlled at a level less than α).

To see that a unique solution exists for $k \leqslant 1/\alpha$, let

$$f(x) = \frac{1}{k-1}P_{2:2}(T_{(2)} \geqslant x,\ T_{(1)} < c_1) + \frac{2}{k}P_{2:2}(T_{(2)} \geqslant x,\ T_{(1)} \geqslant c_1). \tag{8}$$

Now, f is a strictly decreasing function of x with $f(\infty)=0$, so f takes the value α uniquely in the interval $[c_1, \infty)$ unless $f(c_1) < \alpha$. Note that c_1 has been chosen so that $P_{1:1}(T_{(1)} > c_1) = \min\{1/2, k\alpha\}$. By conditioning on the value of T_1, one obtains

$$\begin{aligned} f(c_1) &= \frac{1}{k-1}P_{2:2}(T_1 \geqslant c_1,\ T_2 < c_1) + \frac{1}{k-1}P_{2:2}(T_1 < c_1,\ T_2 \geqslant c_1) \\ &\quad + \frac{2}{k}P_{2:2}(T_1 \geqslant c_1,\ T_2 \geqslant c_1) \\ &= \frac{2}{k(k-1)}[kP_{2:2}(T_1 \geqslant c_1,\ T_2 < c_1) + kP_{2:2}(T_1 \geqslant c_1,\ T_2 \geqslant c_1) \\ &\quad - P_{2:2}(T_1 \geqslant c_1,\ T_2 \geqslant c_1)]. \end{aligned}$$

Now, using the fact that

$$P_{1:1}(T_1 \geqslant c_1) = P_{2:2}(T_1 \geqslant c_1,\ T_2 < c_1) + P_{2:2}(T_1 \geqslant c_1,\ T_2 \geqslant c_1),$$

we obtain

$$\begin{aligned} f(c_1) &= \frac{2}{k(k-1)}[(k-1)P_{1:1}(T_1 \geqslant c_1) + P_{2:2}(T_1 \geqslant c_1,\ T_2 < c_1)] \\ &\geqslant \frac{2}{k}\min\left\{\frac{1}{2}, k\alpha\right\} \\ &= \min\left\{\frac{1}{k}, 2\alpha\right\}. \end{aligned}$$

Thus, $f(c_1)$ exceeds α and a unique solution for $c_2 \in [c_1, \infty)$ with FDR $= \alpha$ exists as long as $k \leqslant 1/\alpha$.

In general, c_j is obtained by considering $c_1, c_2, \ldots, c_{j-1}$ fixed and setting (5) with $m = j$ equal to α,

$$\alpha = \sum_{r=k-j+1}^{k} \frac{r-k+j}{r} \int_0^\infty \int_{-\infty}^\infty P(Z_{(j)} \geqslant d_j,\ Z_{(j-1)} \geqslant d_{j-1}, \ldots, Z_{(k-r+1)} \geqslant d_{k-r+1}, Z_{(k-r)} < d_{k-r}) \phi(z_0)\, \mathrm{d}z_0\, f_v(u)\, \mathrm{d}u. \tag{9}$$

If a solution to (9) exists in $[c_{j-1}, \infty)$, then it is c_j. Otherwise, c_j is set equal to c_{j-1}. We do not know a general formula expressing the condition for which the FDR will equal α, but by construction we have the following main result for the step-down procedure.

Theorem 1. *The step-down procedure using critical constants $\{c_1, c_2, \ldots, c_k\}$ described above, with the test statistics described in Section 2, and with $\hat{\theta}_1, \ldots, \hat{\theta}_k$ consistent estimators of $\theta_1, \ldots, \theta_k$ based on a sample of size N, has*

$$\lim_{N \to \infty} \mathrm{FDR} \leqslant \alpha.$$

In general, the estimates $\hat{\theta}_1, \ldots, \hat{\theta}_k$ may have different sample sizes. In that case the theorem remains true with the limit taking each sample size to infinity. We next consider selecting critical constants for a step-up procedure.

4. Step-up procedure

The majority of the work of this section mirrors that of the previous section. However, there are substantial differences in the calculation of the critical constants for step-down and step-up procedures. We shall use the same notation here as in Section 3, but the step-up critical constants shall be denoted $c_1^*, c_2^*, \ldots, c_k^*$. We start with a basic expression of the FDR,

$$\mathrm{FDR} = \sum_{r=1}^{k} \frac{1}{r} E(V|R=r) \times P_{n:k}(T_{(1)} < c_1^*,\ T_{(2)} < c_2^*, \ldots, T_{(k-r)} < c_{(k-r)}^*,\ T_{(k-r+1)} \geqslant c_{(k-r+1)}^*). \tag{10}$$

From (2), this is asymptotically equivalent (as $N \to \infty$) to

$$\sum_{r=k-m+1}^{k} \frac{(r-k+m)}{r} P_{n:k}(T_{(1)} < c_1^*, \ldots, T_{(k-r)} < c_{k-r}^*, T_{(k-r+1)} \geqslant c_{k-r+1}^*) \approx \sum_{r=k-m+1}^{k} \frac{r-k+m}{r} P_{m:m}(T_{(1)} < c_1^*, \ldots, T_{(k-r)} < c_{k-r}^*,\ T_{(k-r+1)} \geqslant c_{k-r+1}^*). \tag{11}$$

Again, if $c_1^*, c_2^*, \ldots, c_k^*$ are chosen so that (11) is contained below α for any $m \in \{1, 2, \ldots, k\}$, then the FDR will be held below α. Using the same representation as

before, we obtain the following expression for the asymptotic FDR,

$$\sum_{r=k-m+1}^{k} \frac{r-k+m}{r} \times \int_0^\infty \int_{-\infty}^\infty P(Z_{(1)} < d_1^*,\ Z_{(2)} < d_2^*, \ldots, Z_{(k-r)} < d_{k-r}^*,\ Z_{(k-r+1)} \geqslant d_{k-r+1}^*) \times \phi(z_0)\,\mathrm{d}z_0\, f_v(u)\,\mathrm{d}u. \tag{12}$$

where $d_1^*, \ldots, d_k^*$ are defined analogously to the $d_1, \ldots, d_k$ of Section 3. Again by symmetry, the condition in the probability term of (12) that $Z_{(k-r+1)} \geqslant d_{k-r+1}^*$ is equivalent to choosing $m-k+r$ Z's from $Z_1, \ldots, Z_m$ which must be greater than or equal to d_{k-r+1}^*. Thus, we obtain for the probability term in (12)

$$\binom{m}{k-r} P(Z_1 \geqslant d_{k-r+1}^*, \ldots, Z_{m-k+r} \geqslant d_{k-r+1}^*) P(Z_{(1)} < d_1^*, \ldots, Z_{(k-r)} < d_{k-r}^*)$$

$$= \binom{m}{k-r} [1 - \Phi(d_{k-r+1}^*)]^{m-k+r} F_{k-r}(d_1^*, \ldots, d_{k-r}^*), \tag{13}$$

where $F_{k-r}(\cdot, \ldots, \cdot)$ can be evaluated recursively by (7). Expression (12) may now be evaluated using the same numerical integration program used to evaluate (5).

One might try to recursively determine $c_1^*, \ldots, c_k^*$ from (11) with $m = 1, 2, \ldots, k$ as in the step-down case. Here, the difference between step-down and step-up tests becomes critical. The problem encountered is that given a value of c_1^* found from setting (11) with $m = 1$ equal to α, there may not exist a $c_2^* \in [c_1^*, \infty)$ satisfying (11) with $m = 2$ less than or equal to α. One solution is to choose c_1^* to be the maximum from the set of solutions in $[0, \infty)$ to the equations

$$\begin{aligned} \frac{1}{k} P_{1:1}(T_{(1)} \geqslant x) &= \alpha, \\ \frac{2}{k} P_{2:2}(T_{(1)} \geqslant x) &= \alpha f, \\ &\vdots \\ \frac{k}{k} P_{k:k}(T_{(1)} \geqslant x) &= \alpha f, \end{aligned} \tag{14}$$

where f is a chosen fraction between zero and one. The equations in (14) represent the terms that only involve c_1^* from (11) for $m = 1, 2, \ldots, k$ and with x replacing c_1^*. For any equation that does not have a solution in $[0, \infty)$, one can always obtain the maximum of the left-hand side by taking $x = 0$. Furthermore, in this case the value of the left-hand side at $x = 0$ will be less than the right-hand side. Therefore, at the attained maximum solution point of c_1^*, each equation will hold as an inequality with $\leqslant$ replacing $=$. One can then proceed by considering c_1^* as fixed and then setting c_2^* to be the maximum over $[c_1^*, \infty)$ of the solutions to the equations obtained by considering only the terms of (11) for $m = 2, 3, \ldots, k$ which involve only c_1^* and c_2^*, and setting them equal to some value less than α. Although other constructions are possible, we have considered a simple allocation scheme that sets each equation to the sum of the

value attained by the corresponding equation at the previous stage and a fixed fraction, f, of the difference from α. This is made explicit in the general case in which one obtains c_j^* by considering $c_1^*,\ldots,c_{j-1}^*$ as fixed. Define $\beta(j-1,i)$ for $i=j+1,\ldots,k$ to be the value of (11) for $m=i$ with only the terms involving $c_1^*,\ldots,c_{j-1}^*$,

$$\begin{aligned}
&\sum_{r=k-j+2}^{k}\frac{r-k+j+1}{r}P_{j+1:j+1}(T_{(1)}<c_1^*,\ldots,T_{(k-r)}<c_{k-r}^*,T_{(k-r+1)}\geqslant c_{k-r+1}^*)\\
&\quad=\beta(j-1,j+1),\\
&\sum_{r=k-j+2}^{k}\frac{r-k+j+2}{r}P_{j+2:j+2}(T_{(1)}<c_1^*,\ldots,T_{(k-r)}<c_{k-r}^*,T_{(k-r+1)}\geqslant c_{k-r+1}^*)\\
&\quad=\beta(j-1,j+2),\\
&\quad\vdots\\
&\sum_{r=k-j+2}^{k}\frac{r}{r}P_{k:k}(T_{(1)}<c_1^*,\ldots,T_{(k-r)}<c_{k-r}^*,T_{(k-r+1)}\geqslant c_{k-r+1}^*)=\beta(j-1,k). \qquad (15)
\end{aligned}$$

Then c_j^* is found as the maximum of the solutions (in $[c_{j-1}^*,\infty)$) to the equations

$$\begin{aligned}
&\sum_{r=k-j+2}^{k}\frac{r-k+j}{r}P_{j:j}(T_{(1)}<c_1^*,\ldots,T_{(k-r)}<c_{k-r}^*,T_{(k-r+1)}\geqslant c_{k-r+1}^*)\\
&\quad+\frac{1}{k-j+1}P_{j:j}(T_{(1)}<c_1^*,\ldots,T_{(j-1)}<c_{j-1}^*,T_{(j)}\geqslant x)=\alpha,\\
&\sum_{r=k-j+2}^{k}\frac{r-k+j+1}{r}P_{j+1:j+1}(T_{(1)}<c_1^*,\ldots,T_{(k-r)}<c_{k-r}^*,T_{(k-r+1)}\geqslant c_{k-r+1}^*)\\
&\quad+\frac{2}{k-j+1}P_{j+1:j+1}(T_{(1)}<c_1^*,\ldots,T_{(j-1)}<c_{j-1}^*,T_{(j)}\geqslant x)\\
&\qquad=\beta(j-1,j+1)+[\alpha-\beta(j-1,j+1)]f,\\
&\quad\vdots\\
&\sum_{r=k-j+2}^{k}\frac{r}{r}P_{k:k}(T_{(1)}<c_1^*,\ldots,T_{(k-r)}<c_{k-r}^*,T_{(k-r+1)}\geqslant c_{k-r+1}^*)\\
&\quad+P_{k:k}(T_{(1)}<c_1^*,\ldots,T_{(j-1)}<c_{j-1}^*,T_{(j)}\geqslant x)\\
&\qquad=\beta(j-1,k)+[\alpha-\beta(j-1,k)]f, \qquad (16)
\end{aligned}$$

where the value c_{j-1}^* is taken if all the equations have no solution in $[c_{j-1}^*,\infty)$. Other methods of construction could have been used, but we have chosen a simple "spending" rule where at the jth stage we allow at most a fraction f of the remaining FDR to be used under any assumed number of true null hypotheses. By this, we have a family of procedures indexed by f. For f close to 1, the procedure will spend generously at the beginning and therefore have relatively smaller c_1^* than for f small. Consequently, it will be more likely to reject all (or many) of the null hypotheses for a large value of f. For small f, the procedure will save more room for the future and therefore have relatively larger c_1^* than for larger f. Consequently, it will be more likely to reject

at least one null hypothesis with a small value of f. By construction, we have the following main result for the step-up procedure.

Theorem 2. *The step-up procedure using critical constants* $\{c_1^*, c_2^*, \ldots, c_k^*\}$ *described above, with test statistics described in Section* 2, *and with* $\hat{\theta}_1, \ldots, \hat{\theta}_k$ *consistent estimators of* $\theta_1, \ldots, \theta_k$ *based on a sample of size N, has*

$$\lim_{N \to \infty} \mathrm{FDR} \leqslant \alpha.$$

5. Critical constants and power

In this section, we give some examples of the critical constants used by the proposed step-down and step-up procedures, and compare them to those used by the BHSU procedure in the case of normal estimators (Section 2). The power of the procedures is also compared by recording the average proportion of replicates for which each false null hypothesis is rejected out of a large number of simulated replications. The observed FDR is also reported.

Consider first the case of equal correlation with coefficient ρ. Here we restrict the comparison to the case of $k = 5$ and $\alpha = 0.05$. The case of larger values of k and of arbitrary correlation structure is considered later in this section and in Section 6. The critical constants for the one-sided normal theory step-down (NSD), step-up (NSU), and Benjamini and Hochberg's step-up (BHSU) procedures are given in Table 1 for some possible values of ρ and the number of degrees of freedom. Two versions of the step-up procedure are considered, $f = 0.5$ and 0.9. The same comparison is made for two-sided tests in Table 2. The superiority of the normal theory method is apparent when comparing the step-up procedures and looking at c_1. In fact, when the true correlation is zero (which is sufficient for the BHSU to control the FDR), the value of c_1 for either NSU procedure is far smaller than that for the BHSU. Therefore, one expects the NSU procedures to have a higher power when many or all of the null hypotheses are false. A comparison of the NSD procedure critical constants to those for the step-up procedures is difficult because of the differences in structure of step-down and step-up tests. One can see, however, that the NSD procedure in certain cases is more powerful than the NSU procedure because the critical constants are smaller. This implies that the probability of rejecting at least one null hypothesis will be greater with the NSD procedure. Note that the constants $c_1, \ldots, c_5$ for the NSD and NSU procedures can not be used unless $k = 5$ exactly. Furthermore for a problem with k greater than 5, the constants must be computed from scratch starting with c_1. To do otherwise would be invalid because Eqs. (9) and (16) depend on k.

Although a formula for the power could have been derived explicitly, simulation was used to assess the small sample properties of the procedures. The computational efficiency of simulation is so great that we obtain extremely accurate power estimates quite quickly. Moreover, to obtain similar accuracy from numerical integration would require far more computation. The data were simulated from the central five

Table 1
Critical constants c_j for the procedures (one-sided tests)[a]

Procedure	ρ	df	$j=1$	$j=2$	$j=3$	$j=4$	$j=5$
NSD	0	10	0.700	1.416	1.824	2.217	2.721
		20	0.687	1.368	1.736	2.082	2.507
		30	0.683	1.353	1.709	2.040	2.442
	0.1	10	0.700	1.422	1.825	2.211	2.701
		20	0.687	1.373	1.738	2.078	2.495
		30	0.683	1.358	1.711	2.036	2.431
	0.3	10	0.700	1.430	1.824	2.191	2.646
		20	0.687	1.382	1.739	2.063	2.435
		30	0.683	1.367	1.712	2.024	2.396
	0.5	10	0.700	1.433	1.816	2.156	2.562
		20	0.687	1.386	1.733	2.036	2.389
		30	0.683	1.371	1.707	1.998	2.335
NSU, $f=0.5$	0	10	0.716	1.612	1.906	2.268	2.769
		20	0.695	1.555	1.810	2.123	2.540
		30	0.688	1.536	1.779	2.078	2.471
	0.1	10	0.785	1.576	1.925	2.283	2.772
		20	0.761	1.522	1.827	2.136	2.546
		30	0.753	1.505	1.796	2.091	2.477
	0.3	10	0.925	1.499	1.966	2.313	2.774
		20	0.895	1.451	1.862	2.163	2.551
		30	0.885	1.436	1.830	2.117	2.483
	0.5	10	1.073	1.484	1.964	2.344	2.771
		20	1.035	1.418	1.875	2.190	2.548
		30	1.023	1.397	1.847	2.143	2.481
NSU, $f=0.9$	0	10	0.700	1.631	1.910	2.270	2.769
		20	0.687	1.563	1.811	2.123	2.540
		30	0.683	1.542	1.781	2.078	2.471
	0.1	10	0.700	1.688	1.952	2.296	2.780
		20	0.687	1.614	1.848	2.147	2.551
		30	0.683	1.591	1.816	2.101	2.482
	0.3	10	0.700	1.870	2.089	2.397	2.839
		20	0.687	1.772	1.966	2.231	2.600
		30	0.683	1.742	1.928	2.180	2.528
	0.5	10	0.747	2.076	2.359	2.641	3.041
		20	0.728	1.964	2.191	2.423	2.747
		30	0.722	1.929	2.140	2.358	2.660
BHSU	0	10	1.812	1.948	2.120	2.359	2.764
		20	1.725	1.844	1.994	2.197	2.528
		30	1.697	1.812	1.955	2.147	2.457

[a] $k=5$, FDR $\alpha=0.05$.

dimensional multivariate t distribution with $\sigma^2=1$, common correlation $\rho=0, 0.5, 0.7$, 0.9, and $v=10, 30$. Six parameter configurations were considered for each value of ρ and v:

(1) $\theta_i=0$ for all i,

(2) $\theta_1=\theta_2=\theta_3=\theta_4=0$, $\theta_5=2.0$,

Table 2
Critical constants c_j for the procedures (two-sided tests)[a]

Procedure	ρ	df	$j=1$	$j=2$	$j=3$	$j=4$	$j=5$
NSD	0	10	1.221	1.858	2.240	2.615	3.103
		20	1.185	1.765	2.098	2.417	2.819
		30	1.173	1.736	2.054	2.356	2.732
	0.1	10	1.221	1.858	2.239	2.614	3.098
		20	1.185	1.765	2.098	2.416	2.816
		30	1.173	1.736	2.054	2.355	2.729
	0.3	10	1.221	1.859	2.237	2.601	3.063
		20	1.185	1.766	2.096	2.407	2.790
		30	1.173	1.737	2.053	2.347	2.706
	0.5	10	1.221	1.859	2.230	2.572	2.990
		20	1.185	1.768	2.092	2.385	2.735
		30	1.173	1.739	2.050	2.327	2.657
NSU, $f=0.5$	0	10	1.286	1.999	2.330	2.681	3.169
		20	1.214	1.914	2.169	2.462	2.858
		30	1.192	1.887	2.120	2.396	2.764
	0.1	10	1.292	1.995	2.331	2.681	3.168
		20	1.220	1.910	2.170	2.463	2.857
		30	1.198	1.882	2.121	2.396	2.763
	0.3	10	1.342	1.959	2.342	2.689	3.158
		20	1.268	1.875	2.180	2.471	2.853
		30	1.245	1.848	2.131	2.405	2.761
	0.5	10	1.449	1.892	2.377	2.714	3.150
		20	1.373	1.810	2.207	2.491	2.847
		30	1.349	1.784	2.156	2.424	2.756
NSU, $f=0.9$	0	10	1.221	2.090	2.349	2.690	3.174
		20	1.185	1.953	2.178	2.466	2.859
		30	1.173	1.912	2.126	2.398	2.764
	0.1	10	1.221	2.094	2.352	2.692	3.173
		20	1.185	1.956	2.180	2.467	2.859
		30	1.173	1.914	2.128	2.399	2.764
	0.3	10	1.221	2.132	2.384	2.713	3.174
		20	1.185	1.985	2.206	2.484	2.861
		30	1.173	1.940	2.152	2.416	2.768
	0.5	10	1.221	2.261	2.500	2.807	3.231
		20	1.185	2.082	2.295	2.554	2.897
		30	1.173	2.030	2.234	2.479	2.799
BHSU	0	10	2.228	2.359	2.527	2.764	3.169
		20	2.086	2.197	2.336	2.528	2.845
		30	2.042	2.147	2.278	2.457	2.750

[a] $k=5$, FDR $\alpha=0.05$.

(3) $\theta_1=\theta_2=\theta_3=0,\ \theta_4=\theta_5=2.0,$
(4) $\theta_1=\theta_2=0,\ \theta_3=\theta_4=\theta_5=2.0,$
(5) $\theta_1=\theta_2=\theta_3=0,\ \theta_4=0.5,\ \theta_5=1.0,$
(6) $\theta_1=0,\ \theta_2=0.5,\ \theta_3=1.0,\ \theta_4=1.5,\ \theta_5=2.0,$

Table 3
Observed FDR when $k = 5$ (one-sided tests)[a]

ρ	Configuration #	BHSU	NSD	NSU, $f = 0.5$	DTSU
0	1	0.0497	0.0502	0.0501	0.0502
	2	0.0392	0.0407	0.0455	0.0333
	3	0.0292	0.0336	0.0422	0.0218
	4	0.0195	0.0277	0.0407	0.0135
	5	0.0295	0.0294	0.0327	0.0282
	6	0.0098	0.0115	0.0194	0.0075
0.5	1	0.0432	0.0503	0.0502	0.0408
	2	0.0360	0.0464	0.0489	0.0250
	3	0.0279	0.0446	0.0480	0.0177
	4	0.0190	0.0426	0.0451	0.0132
	5	0.0256	0.0315	0.0371	0.0250
	6	0.0091	0.0191	0.0216	0.0073
0.7	1	0.0378	0.0502	0.0501	0.0501
	2	0.0340	0.0491	0.0498	0.0305
	3	0.0274	0.0487	0.0490	0.0215
	4	0.0191	0.0479	0.0405	0.0155
	5	0.0243	0.0348	0.0395	0.0240
	6	0.0092	0.0245	0.0206	0.0082
0.9	1	0.0328	0.0500	0.0501	0.0505
	2	0.0308	0.0500	0.0458	0.0343
	3	0.0259	0.0499	0.0376	0.0250
	4	0.0188	0.0498	0.0256	0.0173
	5	0.0246	0.0425	0.0357	0.0253
	6	0.0097	0.0336	0.0123	0.0095

[a]FDR $\alpha = 0.05$, $v = 30$.

In each simulation, $\alpha = 0.05$ and the number of replications was $1,000,000$, ensuring a simulation error of at most 0.0005. Table 3 reports the observed FDR for each method when $v = 30$. The results for $v = 10$ were similar. The step-up normal theory method (DTSU) of FWER control due to Dunnett and Tamhane (1992) is included in the tables for comparison.

Table 4 reports the average power for the procedures against several possible alternatives. For ease of reporting, we have averaged the number of rejections of each false hypothesis and reported the results as an average power. Each false hypothesis was given a θ value of 2.0, so we have reported only the number of false null hypotheses.

The observed FDR for the procedures is in all cases appropriately controlled at or below the specified level of 0.05. In fact the maximum FDR observed in any configuration was 0.0503 for both the NSD and NSU with $f = 0.5$. This supports the use of the critical constants determined here, which were generated from an asymptotic analysis. Also, notice that the FDR is always highest for the case where all null hypotheses are true. This is also the case when $v = 10$ (not shown). It appears that the maximum FDR is obtained for $\theta_i = 0$ for all i, but we have not been able to proven this assertion. This observed phenomenon makes sense intuitively since when all the null

Table 4
Average power of the procedures (one-sided tests)[a]

ρ	# of false null hyp.	BHSU	NSD	NSU, $f = 0.5$	DTSU
0	1	0.3477	0.3501	0.3501	0.3455
	2	0.3957	0.4004	0.4216	0.3581
	3	0.4403	0.4586	0.5060	0.3743
	5	0.5207	0.6204	0.7904	0.4399
0.5	1	0.3441	0.3843	0.3396	0.3449
	2	0.3985	0.4327	0.4022	0.3598
	3	0.4440	0.4825	0.4736	0.3789
	5	0.5185	0.5971	0.6899	0.4601
0.7	1	0.3406	0.4190	0.3375	0.4169
	2	0.3996	0.4625	0.3991	0.4281
	3	0.4471	0.5056	0.4904	0.4424
	5	0.5218	0.5970	0.6632	0.5050

[a] $k = 5$, FDR $\alpha = 0.05$, $v = 30$.

Table 5
Observed FDR when $k > 5$ (one-sided tests)[a]

k	ρ	NSD
10	0.0	0.0500
	0.7	0.0517
20	0.0	0.0499
	0.7	0.0499
30	0.0	0.0495
	0.7	0.0486

[a] FDR $\alpha = 0.05$, $v = 30$.

hypotheses are true there are the most chances for incorrect rejections, and there are no safe rejections that might bring the FDR down by inflating the denominator, or R (see the discussion in Section 1). The simulated FDR for larger values of k is reported for the NSD method in Table 5. This confirms the usefulness (for larger values of k) of the critical constants derived asymptotically. Since it was observed with $k = 5$ that the FDR is largest when all null hypotheses are true, we have only presented this case in Table 5.

The power of the FDR controlling procedures is higher than the FWER controlling procedure in most cases, as expected. Note that the increase in power for the FDR controlling procedures is greatest when all the null hypotheses are false. The power of the NSD and NSU procedures is higher in all cases than the power of the BHSU procedure, confirming that incorporating the correlation structure raises the power. As in the FWER control case (see Dunnett and Tamhane, 1992), the step-down procedure is more powerful than the step-up procedure when fewer null hypotheses are false. The opposite occurs when more null hypotheses are false.

6. Simulated critical constants

So far we have assumed that the estimators $\hat{\theta}_i$ share a common correlation coefficient. In practice, this is rarely the case even approximately. Here we consider calculation of the critical constants by simulation, without any assumption on the known correlation structure. This is analogous to Dunnett and Tamhane (1995), which extends the FWER controlling procedure for common correlation to the case of arbitrary correlation. We shall use the case of the NSD procedure as an example, although the method can be modified to apply to the NSU procedure as well. Suppose $\hat{\theta}_1,\ldots,\hat{\theta}_k$ have correlation matrix Λ. Eq. (4) is still a valid expression for the asymptotic FDR, where $T_1,\ldots,T_k$ come from the central multivariate t distribution with correlation Λ. Without loss of generality, suppose that the hypotheses have been ordered so that $t_1 \geqslant t_2 \geqslant \cdots \geqslant t_k$, where now the correlation is Λ'. The first constant, c_1, may be found directly from the same equation used in Section 3,

$$\alpha = \frac{1}{k} P_{1:1}(T_{(1)} \geqslant c_1).$$

The second constant, c_2, then is found by considering (4) with $m=2$ set equal to α,

$$\alpha = \frac{1}{k-1} P_{2:2}(T_{(2)} \geqslant c_2,\ T_{(1)} < c_1) + \frac{2}{k} P_{2:2}(T_{(2)} \geqslant c_2,\ T_{(1)} \geqslant c_1). \tag{17}$$

To solve (17), simulate a large number M of k-variate central t statistics $(T_1^*,\ldots,T_k^*)$ with correlation Λ'. For each simulation, order the t statistics, T_1^* and T_2^*, so that $T_{(1)}^* \leqslant T_{(2)}^*$. Next assign each simulation a coefficient, either $1/(k-1)$ or $2/k$, as $T_{(1)}^* < c_1$ or $T_{(1)}^* \geqslant c_1$ respectively. Then order the simulations based on the value of $T_{(2)}^*$. Now consider the sequence of partial sums of the simulation coefficients, starting with the simulation that has the largest $T_{(2)}^*$. Find the simulation for which this partial sum is largest yet still less than or equal to $M\alpha$. The corresponding value of $T_{(2)}^*$ is then taken as c_2, assuming that this is greater than c_1.

The constants are determined recursively, as in Section 3, with the jth constant being determined from the equation

$$\alpha = \sum_{r=k-j+1}^{k} \frac{(r-k+j)}{r} P_{j:j}(T_{(j)} \geqslant c_j,\ldots,T_{(k-r+1)} \geqslant c_{k-r+1}, T_{(k-r)} < c_{k-r}). \tag{18}$$

Eq. (18) is solved from the same set of M simulated k-variate t statistics that were used to determine $c_2,\ldots,c_{j-1}$. The statistics are ordered within each simulation, and the appropriate coefficients assigned to each simulation based on the relative size of $T_{(1)}^*,\ldots,T_{(j-1)}^*$ and $c_1,\ldots,c_{j-1}$. Next the simulations are ordered by the value of $T_{(j)}^*$. The sequence of partial sums of the simulation coefficients are computed, starting with the simulation with the largest $T_{(j)}^*$. Finally, c_j is taken as the value of $T_{(j)}^*$ from the simulation for which the partial sum is greatest but not more than $M\alpha$. Monotonicity is enforced so that $c_j \geqslant c_{j-1}$.

The above algorithm was used to calculate critical constants for the NSD procedure, and an analogous one was implemented for the NSU ($f = 0.5$) procedure. The first

Table 6
Critical constants c_j from simulation with equal correlation (one-sided tests)[a]

Procedure	ρ	$j=1$	$j=2$	$j=3$	$j=4$	$j=5$
NSD	0	0.687	1.363	1.732	2.081	2.503
	0.5	0.687	1.380	1.732	2.031	2.382
NSU, $f=0.5$	0	0.691	1.552	1.801	2.120	2.537
	0.5	1.031	1.410	1.876	2.193	2.546

[a] $k=5$, FDR $\alpha=0.05$, $v=20$.

cases considered are for $k=5$, $\alpha=0.05$, $v=20$, $M=100000$, and one-sided tests. The results appear in Table 6, and serve as a check against Table 1.

Tables of critical constants for the NSD procedure (one-sided or two-sided tests, obtained with $M=50000$) for $k=10(10)30$, $\rho=0$, $0.1(0.2)0.7$, $v=10(10)30$, and $\alpha=0.05$ are available upon request from the author. The FORTRAN program used to generate the constants for any specified correlation matrix is also available. Each particular correlation structure can be tested empirically by simulation to verify the small sample properties before being used in practice.

7. Adjusted FDR-values

Following Westfall and Young (1989,1993), Dunnett and Tamhane (1992), and Troendle (1996), we define the adjusted p-value for a multiple comparison procedure on a given hypothesis as the smallest overall error rate at which the hypothesis can be rejected when the procedure is used on the observed test statistics. Although the term "adjusted p-value" is used for procedures that control the FWER, it can be used just as easily for FDR controlling procedures. One may prefer to call the resulting value the "adjusted FDR-value". Yekutiely and Benjamini (1999) define "FDR p-value correction", which is related to the adjusted FDR-value given here. For adjusted FDR-value to have meaning, we should like to have that any time a given hypothesis is rejected at FDR-level α_0, then it is also rejected by the procedure at any FDR-level $\alpha_1 > \alpha_0$. For the stepwise procedures NSD and NSU, this means that

$$c_j(\alpha_1) \leqslant c_j(\alpha_0), \quad j=1,2,\ldots,k, \tag{19}$$

where $c_j(\alpha)$ is the jth critical constant for the procedure at FDR-level α. For the NSD procedure, inequality (19) is true for $j=1$ by definition, and is shown to hold for $j=2$ in the appendix. We conjecture that (19) holds for all j for both the NSD and NSU ($f=0.5$) procedures. The critical constants have been calculated for the FDR-levels $0.01, 0.50(0.01)$ under the equicorrelated $\rho=0$ and 0.5 cases with $k=5$ and $v=20$. The constants, $c_j(\alpha)$, for the NSD procedure are graphed for $\rho=0$ in Fig. 1. The other graphs look similar. The constants appear to be monotone.

In the FWER control case there is a concern related to (19). The question faced by Dunnett and Tamhane is as follows. For a fixed $\alpha \in (0,1)$, does there exist a

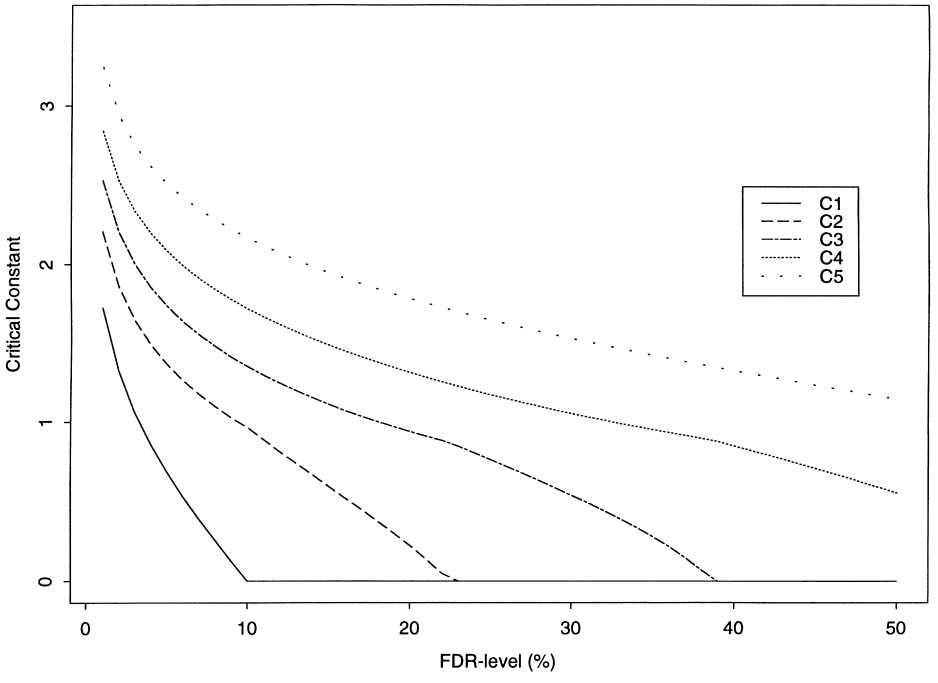

Fig. 1. Critical constants for NSD procedure with $\rho = 0$ and $v = 20$ (one-sided tests).

non-decreasing sequence of constants c_j such that

$$G_k(c_1, \ldots, c_k) = 1 - \alpha, \tag{20}$$

where $G_k(\cdot, \ldots, \cdot)$ denotes the joint distribution function of the order statistics of a multivariate t vector? Notice that if such constants c_j exist, then the FWER of the corresponding step-up procedure will be exactly α. Finner et al. (1993) show that in the independent case the existence of monotone constants satisfying (20) is a distribution free question. Dalal and Mallows (1992) prove that monotone constants satisfying (20) exist in the independent case. For the FDR control case considered here, we force (by construction) the constants c_j to be monotone. Thus, the question of existence of monotone constants is replaced by (19). As a consequence of the forced monotonicity of the c_j, the FDR of the procedures derived here is not held at α, but only less than or equal to α.

Assuming that the critical constants satisfy (19), then adjusted FDR-values can be obtained by comparing the observed test statistics with the critical constants from a sequence of FDR-levels. In other words, one can obtain adjusted FDR-values to a desired accuracy by interval bisection on α by evaluating the critical constants at the given level and comparing to the observed test statistics.

Some examples of adjusted FDR-values obtained for the NSD and NSU ($f = 0.5$) procedures are given for $k = 5$ in Table 7. The artificial examples are analyzed to

Table 7
Adjusted FDR-values for example data (one-sided tests)[a]

Hypothesis #	1	2	3	4	5
Test Statistic	0.6	1.0	1.1	1.5	2.6
P-value	0.278	0.165	0.142	0.075	0.009
Adj. FDR for NSD	0.145	0.145	0.145	0.129	0.033
Adj. FDR for NSU, $f = 0.5$	0.112	0.110	0.110	0.110	0.045
Test Statistic	−0.6	0.3	0.6	1.5	2.0
P-value	0.722	0.384	0.278	0.075	0.030
Adj. FDR for NSD	1.000	0.258	0.258	0.129	0.102
Adj. FDR for NSU, $f = 0.5$	1.000	0.306	0.306	0.195	0.187
Test Statistic	−0.6	1.5	2.0	2.5	2.6
P-value	0.722	0.075	0.030	0.011	0.009
Adj. FDR for NSD	1.000	0.041	0.033	0.033	0.033
Adj. FDR for NSU, $f = 0.5$	1.000	0.042	0.040	0.025	0.025

[a] $\rho = 0.5,\ v = 20$.

provide some comparison of the methods. The adjusted FDR-values were determined to within 0.0001 of the true value and one-sided testing was used. The correlations were assumed equal at $\rho = 0.5$ and $v = 20$. The results confirm the suspicion that the step-down procedure will give smaller adjusted FDR-values when one test statistic dominates the rest. Also, the step-up procedure gives smaller adjusted FDR-values when there are several large test statistics.

8. Two-sided procedures

The theory for two-sided tests is basically the same as that for one-sided tests given in Sections 3 and 4. Here, we give a brief description of the differences. The computational algorithms are straightforward extensions of those for the one-sided tests.

For the step-down procedure NSD, we start with the basic expression of the limiting FDR. In this case it is the same as (4) except that the $T_{(i)}$ are replaced by $|T|_{(i)}$ and the ordering is established by the relative sizes of the absolute value of the test statistics instead of by the signed values. Let d_i and e_i be defined as

$$d_i = \frac{c_i u + \sqrt{\rho} z_0}{\sqrt{1-\rho}}, \qquad e_i = \frac{-c_i u + \sqrt{\rho} z_0}{\sqrt{1-\rho}}, \quad 1 \leqslant i \leqslant k.$$

Then, after representing with the Z_i from Section 3, one sees that the events $Z_{(i)} \geqslant d_i$ obtained in Section 3 are now $(Z_{(i)} \geqslant d_i) \cup (Z_{(i)} \leqslant e_i)$. Thus, in place of (6), we have

$$\binom{m}{k-r} [\Phi(d_{k-r}) - \Phi(e_{k-r})]^{k-r} \times P((Z_{(m-k+r)} \geqslant d_m) \cup (Z_{(m-k+r)} \leqslant e_m) \cap \cdots \cap (Z_{(1)} \geqslant d_{k-r+1}) \cup (Z_{(1)} \leqslant e_{k-r+1})), \tag{21}$$

which can be computed by an algorithm analogous to (7).

In a similar manner we obtain the two-sided testing version of (13) for the NSU procedure,

$$\binom{m}{k-r}[1-\Phi(d_{k-r+1})+\Phi(e_{k-r+1})]^{m-k+r}$$
$$P((Z_{(1)}\leqslant d_1)\cap(Z_{(1)}\geqslant e_1)\cap\cdots\cap(Z_{(k-r)}\leqslant d_{k-r})\cap(Z_{(k-r)}\geqslant e_{k-r})), \tag{22}$$

which can be computed by an algorithm analogous to (7).

One further note about the constants for two-sided tests is needed. In Sections 3 and 4, for ease of computation, we use the convention that $c_1 \in [0,\infty)$. To keep consistent with that decision, we use the restriction $c_1 \in [t_v(0.25),\infty)$ for the two-sided tests. Thus, in either case the unadjusted p-value must be less than 0.5 for there to be any chance that the corresponding null hypothesis will be rejected.

9. Discussion

We have described two normal theory methods of testing multiple hypotheses that control the FDR. The test statistics must be multivariate t statistics which most naturally would arise in the linear model setup, although it could be used in the multiple outcome setting as well if the variances of the outcomes are equal and the correlations estimated. The methods have been shown to asymptotically control the FDR, and simulations support the notion that the FDR is controlled for small samples as well. The power of the new procedures is seen to be much higher than that of a step-up procedure which controls the FDR when the test statistics are independent. Both the step-down and step-up procedures appear to have favorable power against certain alternatives. Adjusted FDR-values appear meaningful for the procedures, and can be obtained by a computational procedure.

Although the use of the FDR is not appropriate for every analysis, there are cases in which it is desirable to allow false discoveries in order to obtain a less exclusive criterion. If one plans on further testing, then one may use FDR procedures as a way of selecting a smaller set of potentially important members. To obtain the most powerful FDR procedure, one should consider incorporating the correlation in the adjustment of the individual p-values to obtain FDR-values. The methods described here do this through the use of critical constants obtained by numerical integration for small k and simulation for large k.

Appendix A. Proof of (19) for NSD with $j = 2$

Define

$$F(x,y)=\frac{1}{k-1}P_{2:2}(T_{(2)}\geqslant y,\ T_{(1)}<x)+\frac{2}{k}P_{2:2}(T_{(2)}\geqslant y,\ T_{(1)}\geqslant x).$$

Then $F(x, y)$ is a monotone decreasing function of y. Now suppose that $c_2(\alpha_1) > c_2(\alpha_0)$. Then we have

$$\begin{aligned}\alpha_1 - \alpha_0 &= F(c_1(\alpha_1), c_2(\alpha_1)) - F(c_1(\alpha_0), c_2(\alpha_0)) \\ &\leqslant F(c_1(\alpha_1), c_2(\alpha_0)) - F(c_1(\alpha_0), c_2(\alpha_0)). \qquad \text{(A.1)}\end{aligned}$$

After combining terms and simplifying (A.1) becomes

$$\alpha_1 - \alpha_0 \leqslant \frac{k-2}{k(k-1)} P_{2:2}(T_{(2)} \geqslant c_2(\alpha_0), c_1(\alpha_1) \leqslant T_{(1)} < c_1(\alpha_0)).$$

Using the fact that T_1 and T_2 are multivariate t statistics, we have that

$$P_{2:2}(T_{(2)} \geqslant c_2(\alpha_0), c_1(\alpha_1) \leqslant T_{(1)} < c_1(\alpha_0)) \leqslant P_{2:2}(c_1(\alpha_1) \leqslant T_1 < c_1(\alpha_0)),$$

and hence we get

$$\begin{aligned}\alpha_1 - \alpha_0 &\leqslant \frac{k-2}{k(k-1)} k(\alpha_1 - \alpha_0) \\ &< \alpha_1 - \alpha_0,\end{aligned}$$

a contradiction, which establishes that $c_2(\alpha_1) \leqslant c_2(\alpha_0)$.

References

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

Benjamini, Y., Yekutiely, D., 1998. The control of the false discovery rate in multiple testing under dependency. unpublished manuscript.

Dalal, S.R., Mallows, C.L., 1992. Buying with exact confidence. Ann. Appl. Probab. 2, 752–765.

Dunnett, C.W., Tamhane, A.C., 1990. A step-up multiple test procedure. Technical Report 90-1, Northwestern University, Dept. of Statistics.

Dunnett, C.W., Tamhane, A.C., 1992. A step-up multiple test procedure. J. Amer. Statist. Assoc. 87, 162–170.

Dunnett, C.W., Tamhane, A.C., 1995. Step-up multiple testing of parameters with unequally correlated estimates. Biometrics 51, 217–227.

Finner, H., Hayter, A.J., Roters, M., 1993. On the joint distribution function of order statistics with reference to step-up multiple test procedures. Technical Report 93-19, University of Trier.

Johnson, N.L., Kotz, S., 1976. Distributions in Statistics: Continuous Multivariate Distributions. Wiley, New York.

Troendle, J.F., 1996. A permutational step-up method of testing multiple outcomes. Biometrics 52, 846–859.

Westfall, P.H., Young, S.S., 1989. P-value adjustments for multiple tests in multivariate binomial models. J. Amer. Statist. Assoc. 84, 780–786.

Westfall, P.H., Young, S.S., 1993. Resampling Based Multiple Testing: Examples and Methods for P-Value Adjustment. Wiley, New York.

Yekutiely, D., Benjamini, Y., 1999. Resampling based false discovery rate controlling multiple test procedures for correlated test statistics. J. Statist. Plann. Inference, in press.