ELSEVIER

Journal of Statistical Planning and Inference 104 (2002) 351–362

journal of statistical planning and inference

www.elsevier.com/locate/jspi

A modified Benjamini–Hochberg multiple comparisons procedure for controlling the false discovery rate

Koon Shing Kwong[a], Burt Holland[b,*], Siu Hung Cheung[c]

[a] *Department of Statistics and Applied Probability, National University of Singapore, 10 Kent Ridge Crescent, Singapore 119260, Singapore*

[b] *Department of Statistics, Temple University, Speakman Hall (006-00), Philadelphia, PA 19122, USA*

[c] *Department of Statistics, Chinese University of Hong Kong, Shatin, NT, Hong Kong*

Received 10 October 2000; received in revised form 21 March 2001; accepted 7 June 2001

Abstract

When performing simultaneous statistical tests, the Type I error concept most commonly controlled by analysts is the familywise error rate, i.e., the probability of committing at least one Type I error. However, this criterion is unduly stringent for some practical situations and therefore may not be appropriate. An alternative concept of error control was provided by Benjamini and Hochberg (J. Roy. Statist. Soc. B 57 (1995) 289) who advocate control of the expected proportion of falsely rejected hypotheses which they term the false discovery rate or FDR. These authors devised a step-up procedure for controlling the FDR. In this article, when the joint distribution of test statistics is known, continuous, and positive regression dependent on each one from a subset of true null hypotheses, we derive and discuss a modification of their procedure which affords increased power. An example is provided to illustrate our proposed method.

MSC: 62F03; 62H15

Keywords: Familywise error rate; FDR; MBH; Positive regression dependency

1. Introduction

Analysts are increasingly confronting large data-mining environments due to advances in computing facilities and data collection strategies. Accompanying such environments in many scientific areas (for example, physiology and agrobiology) are increases in the demand for simultaneous testing techniques. When simultaneous testing of a family of related hypotheses is required, multiple comparison procedures (MCPs) are useful tools. Such procedures provide efficient means to examine individual hypotheses while at the same time controlling some form of overall error probability at a designated level. Most investigators should be aware of the multiplicity problem and

* Corresponding author. Tel.: +1-215-204-8161.

E-mail address: bholland@surfer.sbm.temple.edu (B. Holland).

0378-3758/02/$ - see front matter
PII: S0378-3758(01)00252-X

the desirability of controlling for it, for as explained by Tukey (1977), the failure to control the overall error rate when a large data set undergoes extensive data splitting will inflate the possibility of 'false significance'.

Traditionally, when tackling the multiplicity problem, many researchers have preferred to control the familywise Type I error rate (FWE), i.e., the probability of rejecting at least one true null hypothesis. However, FWE procedures are notorious for their lack of power to detect real differences when the family size is large. Furthermore, as argued by Benjamini and Hochberg (1995), controlling FWE in multiple testing settings may not always be appropriate, especially when the experiment is of an exploratory nature. In such cases, FWE appears to be too stringent a form of simultaneous error control.

As an alternative to FWE control, Benjamini and Hochberg (1995) proposed the false discovery rate (FDR) as a criterion for overall Type I error control. FDR is defined as the expected proportion of incorrect rejections $E(V/R)$, where V is the number of rejected true null hypotheses and R is the number of rejected null hypotheses. When R is 0, V/R is defined to be 0.

Several of our references discuss situations where controlling FDR may be a reasonable or attractive alternative to controlling FWE. In general, this occurs when an overall conclusion for a family of inferences is not determined by the conclusion for one particular inference in the family. For example, in an exploratory drug screening of several new treatments vs. a standard, it may be worthwhile to detect all viable treatments at the risk of misidentifying one or a few treatments as promising.

Based on the idea of FDR control, Benjamini and Hochberg (1995) developed a multiple comparison procedure which controls FDR at a designated value, say α. We refer to this procedure as *BH*. It has been shown that *BH* has excellent power to detect real differences as compared to common FWE-controlling procedures, especially in large-scale multiple testing environments; see Benjamini and Hochberg (1995), Drigalenko and Elston (1997) and Williams et al. (1999). In this paper, we derive a FDR controlling procedure that is more powerful than *BH* when the joint distribution of the test statistics is known and belongs to an important class of distributions (positive regressive dependent on each one from a subset of true null hypotheses). *BH* is briefly reviewed in Section 2. Section 3 introduces the improved modification of *BH*. The determination of the parameter needed to implement this modification is discussed in Section 4. The new procedure is illustrated with an example in Section 5.

2. *BH* procedure and some existing extensions

Assume that there are m null hypotheses $\{H_1, H_2, \ldots, H_m\}$ to be tested and let the ordered p-values be $P_{(1)} \leqslant P_{(2)} \leqslant \cdots \leqslant P_{(m)}$. Denote the null hypothesis associated with $P_{(k)}$ by $H_{(k)}$ for $k = 1, \ldots, m$. *BH* is as follows:

BH procedure

For a given $0 < \alpha < 1$, let

$$k_0 = \max\{k: P_{(k)} \leqslant q_k\},$$

where $q_k = (k/m)\alpha$. Then reject $H_{(1)}, \ldots, H_{(k_0)}$ if k_0 exists, otherwise retain all null hypotheses.

Benjamini and Hochberg (1995) showed that *BH* controls the FDR at level α if the test statistics (or p-values) are independently distributed. Recently, Benjamini and Yekutieli (2001) (called BY hereafter) proved that *BH* also controls the FDR if the joint distribution of the test statistics is positive regression dependent on each one (PRDS) from a subset of true null hypotheses. In an extensive simulation study, Benjamini et al. (1999) also found that *BH* controlled FDR in the family of all pairwise comparisons of means in a balanced one-way layout with normal errors.

An adaptive procedure developed from *BH* was proposed by Benjamini and Hochberg (2000) who utilized the data to estimate the number of true null hypotheses in the family. This procedure was shown to improve the power substantively, but unfortunately, a solid theoretical foundation regarding the control of FDR is not available even for the case where the statistics are independent.

It is important to note that the implementation of the above procedures only uses the marginal distribution of the test statistics, hence paying no attention to their dependency structure. Several attempts have been made to improve *BH* by incorporating the information of the dependency structure of the test statistics into the procedure.

Along this methodological path, the first attempt was provided by Yekutieli and Benjamini (1999) who applied resampling technique to improve *BH*. Then Troendle (2000) successfully derived another powerful procedure which assumes normality of the test statistics. His procedure was proved to control FDR asymptotically.

Next we define the concept of PRDS; a more comprehensive exposition of the idea of PRDS appears in BY. Let I denote the index set $\{1, \ldots, m\}$ and I_0 be a subset of I. Let $\mathbf{y} = (y_1, \ldots, y_m)'$ and $\mathbf{z} = (z_1, \ldots, z_m)'$ be $m \times 1$ vectors. Define an increasing set D such that if an $m \times 1$ vector $\mathbf{y} \in D$ and $z_i \geqslant y_i$ for $i = 1, \ldots, m$, then $\mathbf{z} \in D$. Let $\mathbf{Y}$ be a random vector. If, for any increasing set D and $i \in I_0$, $\Pr(\mathbf{Y} \in D | Y_i \leqslant y_i)$ is non-decreasing in y_i, then the distribution of $\mathbf{Y}$ is PRDS.

The class of PRDS statistics covers a vast number of practical applications such as screening orthogonal contrasts in a balanced design and many-to-one comparisons in clinical trials. See BY for more examples. In particular, since PRDS includes multivariate total positivity of order 2 (MTP_2) and Studentized multivariate normal distributions, all practical applications provided in Sarkar and Chang (1997) and Dunnett and Tamhane (1992) are covered by our proposed method in this paper.

Note that a joint distribution having the PRDS property is invariant to any monotonic transformation (Eaton, 1986). Hence, if a joint distribution of test statistics has the PRDS property, the joint distribution of the corresponding p-values is also PRDS.

3. A modification of *BH*

In this article, we seek to develop an improved version of *BH*, incorporating the known distribution and dependency structure of the test statistics into the procedure. Our approach has at least three major differences as compared to Troendle (2000). First, our procedure covers a larger class (PRDS) of statistics than the multivariate normal distribution. Second, the assertion that our procedure controls FDR is proved to be true regardless of sample size. Third, it is easier to implement our procedure than Troendle's (2000) procedure using either numerical methods or simulation.

The modified *BH* procedure (*MBH*) discussed in this paper retains the flavour of *BH*. In essence, *MBH* has the same layout as *BH* except that the set of constants $\mathbf{q} = \{q_1, \ldots, q_m\}$ is replaced by another set of constants such that the new procedure is more powerful and at the same time, the FDR is still maintained at level α. In terms of distributional requirement, *MBH* is designed only for the class of PRDS test statistics.

We introduce some notation borrowed from BY. Let m_0 be the number of true hypotheses among the m hypotheses. Without loss of generality, assume that $H_1, \ldots, H_{m_0}$ are true. Let P_i^* be the p-value under H_i for $i = 1, \ldots, m_0$, and let $C_k^{(i)}(\mathbf{q})$ denote the event that exactly k hypotheses are rejected under the set of critical constants $\mathbf{q}$ if H_i for $i \leqslant m_0$ with corresponding p-value P_i^* is one of the k rejected hypotheses. Also let $P_{(1)}^* \leqslant P_{(2)}^* \leqslant \cdots \leqslant P_{(m_0)}^*$ be the ordered sequence of the $\{P_i^*\}$.

MBH relies on the following inequality given in BY. When $m_0 = m$ and given that the joint distribution $\mathbf{F}$ of the test statistics is known and continuous with the PRDS property,

$$1 - \Pr_{\mathbf{F}}\left(\bigcap_{k=1}^{m} \{P_{(k)}^* > q_k\}\right) \leqslant \alpha. \tag{3.1}$$

Since the left-hand side of inequality (3.1) is an increasing function of each q_k, there exists a constant $s \geqslant 1$ such that

$$1 - \Pr_{\mathbf{F}}\left(\bigcap_{k=1}^{m-1} \{P_{(k)}^* > q_k\} \cap \{P_{(m)}^* > s q_m\}\right) = \alpha. \tag{3.2}$$

Define a new set of critical values $\mathbf{q}^* = \{q_1^*, \ldots, q_m^*\}$ with $q_k^* = q_k$ for $k = 1, \ldots, m-1$ and $q_m^* = r^* q_m = r^* \alpha$, and r^* is the largest value of r which satisfies the following two constraints:

(a) With s evaluated in Eq. (3.2),

$$1 \leqslant r \leqslant \min\{s, m\}. \tag{3.3}$$

(b) For $m_0 = 2, \ldots, m-1$,

$$\frac{r-1}{r} \max_{\Omega}\{\Pr_{\mathbf{F}}(P_{(m_0)}^* \leqslant r\alpha \mid P_1^* = P_{l_1}, \ldots, P_{m_0}^* = P_{l_{m_0}})\} \leqslant \alpha\left(\frac{m - m_0}{m_0}\right), \tag{3.4}$$

where the maximum is taken over Ω, defined as all distinct subsets $\{l_1, \ldots, l_{m_0}\} \subset \{1, \ldots, m\}$ with cardinality m_0. In some cases, if we do not want to reject a hypothesis

with individual p-value less than a certain value, say c for $\alpha < c < 1$, we can modify the constraint (3.3) to $1 \leqslant r \leqslant \min(s, m, c/\alpha)$. Note that the left-hand side of inequality (3.4) is an increasing function of r for $r \geqslant 1$ and is equal to 0 for $r = 1$, while the right-hand side is always positive. Therefore, there exist some constants $r \geqslant 1$ satisfying the inequality (3.4) for any given α, m_0 and m.

Now, we propose *MBH* as follows:

MBH procedure

Let

$$k_1 = \max\{k: P_{(k)} \leqslant q_k^*\}.$$

Then reject $H_{(1)}, \ldots, H_{(k_1)}$ if k_1 exists, otherwise retain all null hypotheses.

Theorem 1. *MBH controls the FDR at α if the joint distribution* $\mathbf{F}$ *of the test statistics is known and continuous with PRDS from a subset of true null hypotheses.*

Proof. Let Q_{m_0} be the FDR of *MBH* when the number of true null hypotheses is m_0. For $m_0 = m$, due to Eq. (3.2) and $r^* \leqslant s$, we have $Q_m \leqslant \alpha$. For $m_0 \leqslant m-1$, BY (replace $\mathbf{q}$ with $\mathbf{q}^*$ in their paper) demonstrated that

$$Q_{m_0} = \sum_{i=1}^{m_0} \sum_{k=1}^{m} \frac{q_k^*}{k} \left[\frac{\Pr_{\mathbf{F}}(\{P_i^* \leqslant q_k^*\} \cap C_k^{(i)}(\mathbf{q}^*)}{\Pr_{\mathbf{F}}(P_i^* \leqslant q_k^*)} \right]. \tag{3.5}$$

After substituting the critical values of *MBH* $q_k^* = q_k$ for $k = 1, \ldots, m-1$ and $q_m^* = r^*\alpha$ into Eq. (3.5), we get

$$\begin{aligned}
Q_{m_0} &= \sum_{i=1}^{m_0} \sum_{k=1}^{m-1} \frac{\alpha}{m} \left[\frac{\Pr_{\mathbf{F}}(\{P_i^* \leqslant q_k^*\} \cap C_k^{(i)}(\mathbf{q}^*))}{\Pr_{\mathbf{F}}(P_i^* \leqslant q_k^*)} \right] \\
&\quad + \sum_{i=1}^{m_0} \frac{r^*\alpha}{m} \left[\frac{\Pr_{\mathbf{F}}(\{P_i^* \leqslant r^*\alpha\} \cap C_m^{(i)}(\mathbf{q}^*))}{\Pr_{\mathbf{F}}(P_i^* \leqslant r^*\alpha)} \right] \\
&= \sum_{i=1}^{m_0} \sum_{k=1}^{m} \frac{\alpha}{m} \left[\frac{\Pr_{\mathbf{F}}(\{P_i^* \leqslant q_k^*\} \cap C_k^{(i)}(\mathbf{q}^*))}{\Pr_{\mathbf{F}}(P_i^* \leqslant q_k^*)} \right] \\
&\quad + \left(\frac{r^*\alpha}{m} - \frac{\alpha}{m} \right) \sum_{i=1}^{m_0} \left[\frac{\Pr_{\mathbf{F}}(\{P_i^* \leqslant r^*\alpha\} \cap C_m^{(i)}(\mathbf{q}^*))}{\Pr_{\mathbf{F}}(P_i^* \leqslant r^*\alpha)} \right] \\
&= \frac{\alpha}{m} \sum_{i=1}^{m_0} \sum_{k=1}^{m} \left[\frac{\Pr_{\mathbf{F}}(\{P_i^* \leqslant q_k^*\} \cap C_k^{(i)}(\mathbf{q}^*))}{\Pr_{\mathbf{F}}(P_i^* \leqslant q_k^*)} \right] \\
&\quad + \left(\frac{r^*\alpha}{m} - \frac{\alpha}{m} \right) \frac{m_0 \Pr_{\mathbf{F}}(P_{(m)} \leqslant r^*\alpha)}{r^*\alpha}.
\end{aligned} \tag{3.6}$$

Since the test statistics have the PRDS property, BY (replace $\mathbf{q}$ with $\mathbf{q}^*$ in their paper) also showed that

$$\sum_{k=1}^{m} \frac{\Pr_{\mathbf{F}}(\{P_i^* \leqslant q_k^*\} \cap C_k^{(i)}(\mathbf{q}^*))}{\Pr_{\mathbf{F}}(P_i^* \leqslant q_k^*)} \leqslant 1.$$

Therefore, we obtain the upper bound of Q_{m_0} from Eq. (3.6) as follows:

$$Q_{m_0} \leqslant \frac{\alpha m_0}{m} + \frac{m_0(r^*-1)}{mr^*} \Pr_{\mathbf{F}}(P_{(m)} \leqslant r^*\alpha).$$

Based on the facts that

$$\Pr_{\mathbf{F}}(P_{(m)} \leqslant r^*\alpha) \leqslant \max_{\Omega}\{\Pr_{\mathbf{F}}(P_{(m_0)}^* \leqslant r^*\alpha \,|\, P_1^* = P_{l_1}, \ldots, P_{m_0}^* = P_{l_{m_0}})\}$$

and r^* satisfies the constraints in inequalities (3.3) and (3.4), we obtain

$$\begin{aligned} Q_{m_0} &\leqslant \frac{\alpha m_0}{m} + \frac{m_0(r^*-1)}{mr^*} \max_{\Omega}\{\Pr_{\mathbf{F}}(P_{(m_0)}^* \leqslant r^*\alpha \,|\, P_1^* = P_{l_1}, \ldots, P_{m_0}^* = P_{l_{m_0}})\} \\ &\leqslant \alpha \end{aligned}$$

for $m_0 = 1, \ldots, m-1$. □

Note that when $r^* = 1$, *MBH* reduces to *BH*.

Theorem 2. *MBH is more powerful than BH if the joint distribution* $\mathbf{F}$ *of the test statistics is known and continuous with PRDS from a subset of true null hypotheses.*

Proof. By comparing the two sets of critical values under these two procedures, we observe that $q_m^* \geqslant q_m$ and $q_k^* = q_k$ for $k = 1, \ldots, m-1$. Hence, $k_1 \geqslant k_0$. Therefore, *MBH* is more powerful than *BH*. □

4. Determination of r^* for *MBH* procedure

To facilitate subsequent discussions, let us define $\mathbf{X} = (X_1, \ldots, X_m)$ to be the test statistics with corresponding p-values $(P_1, \ldots, P_m)$ and m null hypotheses $\{H_1, \ldots, H_m\}$. For m_0 true null hypotheses, let $\mathbf{X}_0 = (X_1^*, \ldots, X_{m_0}^*)$ and $(P_1^*, \ldots, P_{m_0}^*)$ be the corresponding test statistics and p-values, respectively. In fact, $\mathbf{X} = \mathbf{X}_0$ if $m_0 = m$. Denote the order values of $\mathbf{X}_0$ by $X_{(1)}^* \leqslant \cdots \leqslant X_{(m_0)}^*$. Assume that the joint distribution of $\mathbf{X}_0$ is known and continuous with the PRDS property.

The crucial step in implementing *MBH* is the computation of r^*. There are two major steps in evaluating r^*. The first one is to compute s in Eq. (3.2) when α is given.

For any given distribution of $\mathbf{X} = \mathbf{X}_0$, Dunnett and Tamhane's (1995) approach to evaluate the probability involving order statistics of correlated random variables can be extended to compute s in (3.2). However, in general, the extension is very complex

computationally when m is large. Fortunately, if the joint probability density function (pdf) of $\mathbf{X}_0$ has the following form:

$$\int \prod_{i=1}^{m} g_i(x_i;\ z_1,\ldots,z_n)h_1(z_1)\cdots h_n(z_n)\prod_{j=1}^{n} \mathrm{d}z_j, \tag{4.1}$$

where $g_i(x_i;\ z_1,\ldots,z_n)$ is a conditional pdf of $X_i^*|Z_1=z_1,\ldots,Z_n=z_n$ for $i=1,\ldots,m$ and the $h_j(z_j)$ are the mutually independent pdfs of Z_j for $j=1,\ldots,n$, the computation of s can be greatly simplified. Since

$$\mathrm{Pr}_{\mathbf{F}}\left(\bigcap_{k=1}^{m}\{P_{(k)}^* > b_k\}\right) = \mathrm{Pr}_{\mathbf{F}}\left(\bigcap_{k=1}^{m}\{X_{(k)}^* \leqslant c_k\}\right), \tag{4.2}$$

where $c_k = F_i(1-b_{m-k+1})$ for $k=1,\ldots,m$ and $F_i(\cdot)$ is a marginal distribution function of $X_{(i)}^*$ for $i=1,\ldots,m$, we can apply Kwong and Liu's (2000) algorithm to efficiently calculate the probability on the right-hand side of (4.2). Then, the constant s in (3.2) can be evaluated for any given α and m. Besides, if the conditional pdf of $X_i^*|Z_1=z_1,\ldots,Z_n=z_n$ in (4.1) has a common distribution, i.e. $\mathbf{X}_0$ belongs to the class of multivariate n order conditionally independent distributions (MCI_n) which is defined by Kwong (2001), we can significantly reduce the complexity of the numerical evaluation by applying the recursive formula in Theorem 1 of Kwong (2001) to efficiently calculate the probability in (4.2).

If the distribution of $\mathbf{X}_0$ does not have the form (4.1), for evaluation of r^* of *MBH*, we suggest a simulation method that is similar to the algorithm given in Dunnett and Tamhane (1995).

Let N_A be the number of simulations to be conducted and N_{T} be the total number of repetitions of generated $\mathbf{X}_0$ with $m_0=m$ in each simulation in order to obtain one simulated s. For simplicity, always set N_{T} such that $N_0=(1-\alpha)N_{\mathrm{T}}$ is an integer. The simulation procedure is as follows:

1. Set a counter $N_C=0$.
2. Simulate a sample of $\mathbf{X}_0$ with $m_0=m$ based on the given distribution and then obtain the corresponding ordered p-values, $P_{(j)}^*$ for $j=1,\ldots,m$.
3. If $P_{(j)}^* \geqslant q_j$ for $j=1,\ldots,m-1$, then store the value of $P_{(m)}^*$, increase N_C by 1. Return to Step 2.
4. After completing the generation of N_{T} samples, pick the (N_C-N_0+1)th ordered value of stored $P_{(m)}^*$, say t.
5. Estimate s by t/α and return to Step 1.
6. After conducting N_A simulations, the final estimate of s is the average of all the estimates obtained in Step 5.

Note that if the random variable N_C is less than N_0 in Step 4, no reasonable estimate of s exists. However, if N_{T} is set sufficiently large, it is very unlikely that N_C will be less than N_0.

After obtaining s, according to constraint (3.4), the second step is to solve the following set of equations for r:

$$\frac{r-1}{r}\max_{\Omega}\{\Pr_{\mathbf{F}}(P^*_{(m_0)} \leqslant r\alpha \mid P^*_1 = P_{l_1}, \ldots, P^*_{m_0} = P_{l_{m_0}})\} = \alpha\left(\frac{m-m_0}{m_0}\right) \tag{4.3}$$

for $m_0 = 2, \ldots, m-1$ and choose the minimum, say r', out of the $m-2$ obtained values of r. Then, r' is the largest possible value of r satisfying the constraint in inequality (3.4). In order to satisfy the other constraint in inequality (3.3), take r^* of MBH to be $\min\{r', s, m\}$. If $\mathbf{X}_0$ has the form in (4.1), the probability in (4.3) can easily be evaluated by

$$\begin{aligned}&\Pr_{\mathbf{F}}(P^*_{(m_0)} \leqslant r\alpha \mid P^*_1 = P_{l_1}, \ldots, P^*_{m_0} = P_{l_{m_0}})\\ &\quad = \int \prod_{i=1}^{m_0}[1 - G_{l_i}(c_i;\ z_1, \ldots, z_n)]h_1(z_1)\cdots h_n(z_n)\prod_{j=1}^{n} \mathrm{d}z_j\end{aligned} \tag{4.4}$$

where $c_i = F_{l_i}(1 - r\alpha)$, and $F_{l_i}(\cdot)$ and $G_{l_i}(\cdot;\ z_1, \ldots, z_n)$ are the cumulative distribution functions of conditional random variables X_{l_i} and $X_{l_i}|Z_1 = z_1, \ldots, Z_n = z_n$ for $i = 1, \ldots, m_0$, respectively.

Similar to the way of simulating s, we can also simulate r which satisfies equation (4.3) for any given α, m_0 and m. Note that there are

$$M = \binom{m}{m_0}$$

distinct elements in $\Omega = \{\omega_1, \ldots, \omega_M\}$. The algorithm is as follows:

1. Calculate $a = \alpha(m - m_0)/m_0$.
2. Set the counter $j = 1$.
3. For ω_j, simulate a sample of $X_{l_1}, \ldots, X_{l_{m_0}}$ under the true hypotheses and then determine their corresponding largest p-value $P^*_{(m_0)}$. Repeat this step N_{T} times.
4. Order the N_{T} of $P^*_{(m_0)}$, say $p_1 \leqslant p_2 \leqslant \cdots \leqslant p_{N_{\mathrm{T}}}$.
5. Set the counter, $n = 1$.
6. Evaluate $r = p_n/\alpha$.
7. If $(r-1)n/(rN_{\mathrm{T}}) - a < 0$, then increase n by 1 and return to Step 6. Otherwise, let $r_j = p_{n-1}/\alpha$.
8. If $j < M$, set $j = j + 1$ and return to Step 3.
9. After all r_j for $j = 1, \ldots, M$ are found, the estimate of r is the minimum of r_j.

After obtaining all $m-2$ values of r for $m_0 = 2, \ldots, m-1$, respectively, the minimum value of the obtained r would be the estimate of r'.

Note that if $\mathbf{X}$ has equi-correlated structure, i.e. the correlation coefficients of any two X_j are equal to a constant, then $\Pr_{\mathbf{F}}(P^*_{(m_0)} \leqslant r\alpha \mid P^*_1 = P_{l_1}, \ldots, P^*_{m_0} = P_{l_{m_0}})$ is equal to a unique constant for any given subset $\{l_1, \ldots, l_{m_0}\} \in \Omega$. Therefore, it is not necessary to search for the maximum probability in (4.3). As a result, the computation time for evaluating r' is reduced significantly for large m in this case.

However, if $\mathbf{X}$ has an arbitrary correlation structure, we have to evaluate the probability in (4.4) under each of $2^m - m - 2$ correlation combinations before r' can be determined. When m is relatively large, it may take a lot of computation time to determine r' using either the numerical method or the simulation approach. Therefore, we propose use of the modified Bonferroni inequality (see Kwerel, 1975) to provide the upper bound of the probability in (4.4), i.e.

$$\Pr_{\mathbf{F}}(P^*_{(m_0)} \leqslant r\alpha \,|\, P^*_1 = P_{l_1}, \ldots, P^*_{m_0} = P_{l_{m_0}}) \leqslant 1 - \frac{S_1 - (S_2/k)}{k+1}$$

where $S_1 = m_0(1 - r\alpha)$,

$$S_2 = \sum_{l_1 \leqslant l_i < l_j \leqslant l_{m_0}} \Pr_{\mathbf{F}}(\{P^*_1 > r\alpha\} \cap \{P^*_2 > r\alpha\} \,|\, P^*_1 = P_{l_i}, P^*_2 = P_{l_j}),$$

and k is the integer part of $1 + 2(S_2/S_1)$. By using this modified Bonferroni inequality, we can easily evaluate the bounds to approximate the exact values and then obtain the upper bound of the maximum of the probability in (4.3) even for very large m.

5. Example: multiple comparison with a control

To demonstrate the application of *MBH*, consider the case of one-sided multiple hypotheses testing of m treatments with a control (Dunnett, 1955) in a balanced one-way layout. Then under the null hypotheses, the m dependent statistics $\mathbf{X}$ have a multivariate t-distribution with equi-correlated structure and degrees of freedom v which depends on the sample size of the m treatments and the control. BY shows that $\mathbf{X}$ has the PRDS property and Kwong (2001) indicates that $\mathbf{X}$ is also MCI_2, where $g(x_i;\, z_1, z_2)$ for $i = 1, \ldots, m$ and $h_1(z_1)$ are the standard normal density function, and $h_2(z_2)$ is the density function of $\sqrt{\chi^2_v/v}$. Therefore, the parameter of *MBH* can be evaluated accordingly. Table 1 presents the computed r^* of *MBH* in one-sided tests with $m = 3(1)8(2)20$, $\alpha = 0.05$, $v = 20(20)100, \infty$ and the correlation coefficient between any two X_j's $\rho = 0(0.1)0.5$.

It is worth noting that in the cases $\rho = 0$ and $v = \infty$, for any m, the test statistics are mutually independent; therefore, *MBH* reduces to *BH*, i.e. $r^* = 1$. For large values of v and ρ at or near 0, r^* changes smoothly but abruptly as either v or ρ changes. For example, for $m = 3$, $v = \infty$, and $\rho = 0.02(0.02)0.08$, the values of r^* given by the program are 1.200, 1.311, 1.391 and 1.453, respectively, and for $m = 14$ and $\rho = 0$, $r^* = 11.076$ if $v = 100$ while $r^* = 1$ if $v = \infty$. For this reason, interpolation is not recommended for $v \geqslant 100$ and $\rho \leqslant 0.1$. Instead, our program for computing the exact value of r^* should be employed. A small amount of dependency information can be used to substantially improve the *BH* procedure.

For illustration purposes, some parameters of *MBH* are also simulated and presented in Table 2. The simulation results for s are based on $N_A = 20$ and $N_T = 200,000$, while that of r is based on $N_T = 1,000,000$. As indicated in the table, the differences between the exact and simulated r^* are negligible.

Table 1
Parameter r^* for MBH with $\alpha = 0.05$

v	ρ	m											
		3	4	5	6	7	8	10	12	14	16	18	20
20	0.0	1.909	3.376	4.946	6.000	7.000	8.000	9.457	10.511	11.463	12.336	12.926	13.458
	0.1	1.949	3.184	4.451	5.612	6.638	7.534	8.998	10.122	10.768	11.287	11.720	12.087
	0.2	1.985	3.072	4.150	5.132	6.004	6.689	7.496	8.124	8.632	9.055	9.412	9.719
	0.3	2.014	2.989	3.929	4.599	5.003	5.348	5.909	6.353	6.715	7.020	7.281	7.508
	0.4	2.033	2.908	3.449	3.753	4.010	4.226	4.577	4.853	5.080	5.270	5.434	5.577
	0.5	2.033	2.676	2.893	3.062	3.200	3.316	3.501	3.646	3.759	3.853	3.921	4.011
40	0.0	1.621	2.855	4.256	5.596	6.802	7.862	9.457	10.512	11.464	12.338	12.934	13.468
	0.1	1.759	2.846	3.993	5.066	6.032	6.888	8.317	9.447	10.359	11.108	11.733	12.106
	0.2	1.849	2.840	3.839	4.761	5.590	6.328	7.513	8.135	8.638	9.057	9.414	9.722
	0.3	1.917	2.831	3.724	4.538	5.046	5.389	5.944	6.385	6.741	7.044	7.302	7.526
	0.4	1.965	2.807	3.489	3.800	4.056	4.274	4.624	4.901	5.129	5.317	5.479	5.623
	0.5	1.991	2.705	2.928	3.103	3.245	3.364	3.550	3.697	3.815	3.912	3.994	4.067
60	0.0	1.488	2.602	3.916	5.200	6.373	7.417	9.145	10.486	11.465	12.338	12.936	13.471
	0.1	1.683	2.708	3.802	4.836	5.771	6.607	8.013	9.136	10.048	10.803	11.437	11.978
	0.2	1.798	2.750	3.716	4.614	5.423	6.147	7.375	8.138	8.633	9.058	9.414	9.722
	0.3	1.881	2.771	3.645	4.445	5.060	5.402	5.955	6.395	6.749	7.052	7.309	7.532
	0.4	1.940	2.768	3.503	3.815	4.072	4.289	4.640	4.916	5.146	5.332	5.480	5.635
	0.5	1.975	2.714	2.939	3.117	3.260	3.379	3.569	3.714	3.833	3.931	4.014	4.087
80	0.0	1.414	2.450	3.709	4.955	6.105	7.136	8.859	10.210	11.281	12.146	12.856	13.447
	0.1	1.645	2.635	3.700	4.710	5.628	6.451	7.841	8.957	9.868	10.624	11.262	11.808
	0.2	1.772	2.703	3.652	4.535	5.334	6.050	7.268	8.140	8.626	9.058	9.413	9.722
	0.3	1.862	2.740	3.604	4.397	5.068	5.409	5.964	6.401	6.753	7.056	7.312	7.534
	0.4	1.928	2.749	3.509	3.822	4.079	4.297	4.647	4.923	5.151	5.339	5.450	5.642
	0.5	1.967	2.719	2.945	3.124	3.267	3.387	3.577	3.723	3.842	3.940	4.024	4.096
100	0.0	1.353	2.326	3.541	4.759	5.892	6.914	8.635	9.992	11.076	11.954	12.677	13.281
	0.1	1.617	2.584	3.629	4.623	5.530	6.344	7.724	8.835	9.745	10.502	11.142	11.690
	0.2	1.754	2.672	3.610	4.484	5.276	5.987	7.198	8.141	8.643	9.058	9.413	9.722
	0.3	1.850	2.720	3.577	4.365	5.072	5.413	5.968	6.404	6.760	7.058	7.314	7.536
	0.4	1.919	2.736	3.513	3.826	4.084	4.301	4.655	4.928	5.155	5.344	5.508	5.646
	0.5	1.962	2.712	2.949	3.128	3.272	3.392	3.582	3.728	3.842	3.945	4.030	4.102
∞	0.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.1	1.505	2.364	3.317	4.235	5.082	5.849	7.166	8.243	9.137	9.888	10.529	11.083
	0.2	1.684	2.545	3.433	4.266	5.025	5.711	6.890	7.860	8.646	9.059	9.415	9.721
	0.3	1.802	2.638	3.468	4.234	4.927	5.429	5.982	6.416	6.770	7.067	7.322	7.501
	0.4	1.886	2.684	3.451	3.844	4.102	4.320	4.674	4.949	5.173	5.361	5.523	5.662
	0.5	1.941	2.683	2.962	3.144	3.290	3.404	3.602	3.748	3.870	3.969	4.052	4.124

In order to compare the performances of MBH with BH in this example, we simulated average powers of the tests, the proportion of the false hypotheses which are correctly rejected. The configurations and structure of this simulation study are similar to the one given in Dunnett and Tamhane (1992). The number of repetitions in each simulated case is 50,000. The estimated standard errors of the simulated power are less

Table 2
Simulated and exact values of parameter r^* of *MBH*

m	ρ	v	Simulated			Exact
			s	r'	r^*	r^*
4	0.4	20	2.895	3.073	2.895	2.908
8	0.1	100	6.295	7.914	6.295	6.344
10	0.5	40	5.942	3.537	3.537	3.550
12	0.2	60	8.356	8.160	8.160	8.138
16	0.3	80	9.087	7.025	7.025	7.052

Table 3
Simulated average powers of *BH* and *MBH* when $\alpha = 0.05$, $\rho = 0.1$, and $v = 60$[a]

δ	m_0	m				
		4	8	12	16	20
2	0	0.543	0.518	0.508	0.503	0.498
		0.664	0.775	0.824	0.854	0.870
	1	0.499	0.495	0.491	0.488	0.485
		0.522	0.588	0.642	0.680	0.711
	2	0.451	0.466	0.473	0.474	0.474
		0.456	0.503	0.548	0.582	0.614
	3	0.398	0.437	0.450	0.459	0.464
		0.399	0.453	0.492	0.526	0.555
	$m/4$				0.446	0.439
					0.486	0.480
	$m/2$		0.406	0.388	0.375	0.365
			0.413	0.396	0.384	0.373
	$3m/4$		0.335	0.310	0.296	0.279
			0.337	0.312	0.298	0.282
3	0	0.894	0.893	0.893	0.892	0.892
		0.957	0.987	0.994	0.996	0.997
	1	0.867	0.879	0.884	0.886	0.887
		0.875	0.907	0.927	0.937	0.944
	2	0.823	0.863	0.873	0.878	0.882
		0.824	0.873	0.892	0.905	0.916
	3	0.758	0.843	0.862	0.869	0.875
		0.759	0.847	0.872	0.885	0.896
	$m/4$				0.861	0.861
					0.870	0.870
	$m/2$		0.816	0.812	0.812	0.811
			0.817	0.814	0.814	0.813
	$3m/4$		0.736	0.724	0.719	0.716
			0.736	0.725	0.720	0.716

[a]Key: in each cell, the first and second entries are simulated average powers of *BH* and *MBH*, respectively.

than 0.0022. We consider the cases where $m = 4(4)20$, $\alpha = 0.05$, $\rho = 0.1$ and $v = 60$ so all the required parameters r^* of *MBH* can be obtained from Table 1. Table 3 shows the simulated average powers of the two procedures when $\delta = 2, 3$ and the numbers of

true hypotheses are $0(1)3, m/4, m/2, 3m/4$, where δ is the number of standard deviations of the difference between the treatment and control sample means when the hypotheses are false. As expected, MBH is uniformly more powerful than BH and the power improvement is most appreciable when m_0 is small and/or m is large.

Acknowledgements

This paper benefitted from helpful comments by James Troendle and two anonymous referees. The work of Burt Holland was supported in part by a Temple University Research and Study Leave.

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., Hochberg, Y., 2000. On the adaptive control of the false discovery rate in multiple testing with independent statistics. J. Educ. Behav. Statist. 25, 60–83.

Benjamini, Y., Hochberg, Y., Kling, Y., 1999. False discovery rate controlling procedures for pairwise comparisons, submitted for publication.

Benjamini, Y., Yekutieli, D., 2001. The control of the false discovery rate in multiple testing under dependency. Ann. of Statist. 29, in press.

Drigalenko, E.I., Elston, R.C., 1997. False discoveries in genome scanning. Genet. Epidemiol. 14, 779–784.

Dunnett, C.W., 1955. A multiple comparison procedure for comparing several treatments with a control. J. Amer. Statist. Assoc. 50, 1096–1121.

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.

Eaton, M.L., 1986. Lectures on Topics in Probability Inequalities. Center for Mathematical and Computational Sciences CWI Tract 35, Amsterdam.

Kwerel, S.M., 1975. Most stringent bounds on aggregated probabilities of partially specified dependent probabilities systems. J. Amer. Statist. Assoc. 70, 472–479.

Kwong, K.S., 2001. An algorithm for construction of multiple hypothesis testing. Comput. Statist. 16, 165–171.

Kwong, K.S., Liu, W., 2000. Calculation of critical values for Dunnett and Tamhane's step-up multiple test procedure. Statist. Probab. Lett. 49, 411–416.

Sarkar, S.K., Chang, C.K., 1997. The Simes method for multiple hypotheses testing with positively dependent test statistics. J. Amer. Statist. Assoc. 92, 1601–1608.

Troendle, J.F., 2000. Stepwise normal theory multiple test procedures controlling the false discovery rate. J. Statist. Plann. Inference 84, 139–158.

Tukey, J.W., 1977. Some thoughts on clinical trials, especially problems of multiplicity. Science 198, 679–684.

Williams, V.S.L., Jones, L.V., Tukey, J.W., 1999. Controlling error in multiple comparisons, with examples from state-to-state differences in educational achievement. J. Educ. Behav. Statist. 24, 42–69.

Yekutieli, D., Benjamini, Y., 1999. Resampling based false discovery rate controlling multiple test procedures for correlated test statistics. J. Statist. Plann. Inference 82, 171–196.