Biometrika (2014), **101**, 4, *pp.* 899–911
Printed in Great Britain
doi: 10.1093/biomet/asu032
Advance Access publication 24 October 2014

A class of improved hybrid Hochberg–Hommel type step-up multiple test procedures

By JIANGTAO GOU

Department of Statistics, Northwestern University, Evanston, Illinois 60208, U.S.A.
jgou@u.northwestern.edu

AJIT C. TAMHANE

Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, Illinois 60208, U.S.A.
atamhane@northwestern.edu

DONG XI

Statistical Methodology, Novartis Pharmaceuticals Corporation, One Health Plaza, East Hanover, New Jersey 07936, U.S.A.
dong.xi@novartis.com

AND DROR ROM

Prosoft, Inc., 996 Old Eagle School Road #1106, Wayne, Pennsylvania 19087, U.S.A.
d.rom@prosoftsoftware.com

Summary

In this paper we derive a new p-value based multiple testing procedure that improves upon the Hommel procedure by gaining power as well as having a simpler step-up structure similar to the Hochberg procedure. The key to this improvement is that the Hommel procedure can be improved by a consonant procedure. Exact critical constants of this new procedure can be numerically determined. The zeroth-order approximations to the exact critical constants, albeit slightly conservative, are simple to use and need no tabling, and hence are recommended in practice. The proposed procedure is shown to control the familywise error rate under independence among the p-values. Simulations empirically demonstrate familywise error rate control under positive and negative dependence. Power superiority of the proposed procedure over competing ones is also empirically demonstrated. Illustrative examples are given.

Some key words: Adjusted p-value; Closure method; Consonant procedure; Familywise error rate; p-value based multiple test procedure; Stepwise procedure.

1. Introduction

Multiple test procedures based on p-values are widely used because of their simplicity and because they do not require strong distributional assumptions. They are especially useful when different test statistics are employed for different hypotheses and can be transformed to a common

p-value scale. In this paper we propose a new class of powerful step-up multiple test procedures based on p-values.

Consider $n \geqslant 2$ hypotheses, $H_1, \ldots, H_n$, and let $p_1, \ldots, p_n$ denote their p-values. We assume the free-combination condition (Holm, 1979) that any combination of the true and false hypotheses is possible and there are no logical dependencies among the hypotheses. We further assume that the p-values are statistically independent and each p_i is uniformly distributed over [0, 1] under H_i; the dependence case is studied via simulation in the sequel. All procedures in this paper satisfy the strong familywise error rate control requirement

$$\sup_{H(I)} \text{pr}\left\{\text{Reject at least one true } H_i,\ i \in I \mid H(I)\right\} \leqslant \alpha$$

for a specified $\alpha \in (0, 1]$ where the notation $\mid H(I)$ indicates that the probability is computed under $H(I)$ and the supremum is taken over all nonempty intersection hypotheses $H(I) = \bigcap_{i \in I} H_i$ with $I \subseteq \{1, \ldots, n\}$ (Hochberg & Tamhane, 1987).

The most basic among the p-value based procedures is the Bonferroni procedure, which rejects any hypothesis H_i with $p_i \leqslant \alpha/n$. Stepwise procedures are more powerful. They use the ordered p-values, $p_{(1)} \leqslant \cdots \leqslant p_{(n)}$, as the test statistics for the corresponding hypotheses $H_{(1)}, \ldots, H_{(n)}$. A step-down version of the Bonferroni procedure proposed by Holm (1979) begins by testing $H_{(1)}$ and continues testing and rejecting $H_{(i)}$ as long as $p_{(i)} \leqslant \alpha/(n - i + 1)$ $(i = 1, \ldots, n)$. Testing stops by accepting the remaining hypotheses, $H_{(i)}, \ldots, H_{(n)}$, if $p_{(i)} > \alpha/(n - i + 1)$. Hochberg (1988) proposed a step-up procedure, which begins by testing $H_{(n)}$ and continues testing and accepting $H_{(i)}$ as long as $p_{(i)} > \alpha/(n - i + 1)$ $(i = n, \ldots, 1)$. Testing stops by rejecting the remaining hypotheses, $H_{(i)}, \ldots, H_{(1)}$, if $p_{(i)} \leqslant \alpha/(n - i + 1)$. For the same α, the Hochberg procedure rejects all hypotheses rejected by the Holm procedure and possibly more. We will refer to this property as being uniformly more powerful.

The Hommel (1988) procedure does not have the simple stepwise structure of the Holm or the Hochberg procedures but it is uniformly more powerful than them. It is explained in more detail in § 2. In summary, we have the following uniform power dominance relationships among the four procedures:

$$\text{Hommel} > \text{Hochberg} > \text{Holm} > \text{Bonferroni}.$$

The Hommel and Hochberg procedures require the assumption of independence of the p-values since they are based on the Simes (1986) test, which makes this assumption, while the Holm and Bonferroni procedures do not have this restriction. Sarkar & Chang (1997) and Sarkar (1998) have shown that the Simes test is conservative under certain types of positive dependence, so the Hommel and Hochberg procedures are also conservative under those conditions.

In this paper we propose a general step-up procedure, which we refer to as the hybrid procedure since it is a combination of the Hommel and Hochberg procedures, and which improves upon them. Improving the Hochberg procedure is not difficult since it is known to be conservative. In fact, Rom (1990) showed how to compute the exact critical constants for the Hochberg procedure and thus sharpen it.

There is a common perception that the Hommel procedure cannot be improved since it is an exact shortcut to the closed procedure that uses the exact α-level Simes test as a local test for all intersection hypotheses. However, it is not a consonant procedure (Westfall et al., 1999, p. 32) according to the following condition due to Hommel et al. (2007): a closed procedure is said to be consonant (Gabriel, 1969) if, when it rejects any intersection hypothesis $H(I) = \bigcap_{i \in I} H_i$, there exists at least one j such that the procedure rejects all intersections $H(J)$ with $j \in J \subseteq I$.

Romano et al. (2011) have shown that a nonconsonant procedure can be improved by a more powerful consonant procedure if power is defined as the probability of rejecting at least one false elementary hypothesis. We use this result to improve upon the Hommel procedure by using a new local test for intersection hypotheses that satisfies the stronger consonance condition. The proposed procedure also improves upon the Hommel procedure in another way by using a simpler step-up testing algorithm similar to that of the Hochberg procedure. We study in detail a special case of the hybrid procedure, called the hybrid-0 procedure, which is uniformly more powerful than the Hochberg procedure for all n and than the Hommel procedure for $n = 3$ or 4, and is generally, but not uniformly, more powerful than the Hommel procedure for $n \geqslant 5$. The hybrid-0 procedure was proposed by Rom (2013) but he did not prove that it controls the familywise error rate for all n, nor did he consider the general hybrid procedure.

2. Review of existing procedures

In this section we explain how the various stepwise procedures based on p-values are derived using the closure method (Marcus et al., 1976). To apply this method requires that we have available a local α-level test of every intersection hypothesis $H(I) = \bigcap_{i \in I} H_i$ where $I = \{i_1, \ldots, i_m\}$ is a nonempty subset of $\{1, \ldots, n\}$. To avoid double subscripts, we will set $I = \{1, \ldots, m\}$, keeping in mind that these are not necessarily the first m hypotheses. Suppose that for every $m \leqslant n$, there exist critical constants $c_{m1} \geqslant \cdots \geqslant c_{mm}$ such that

$$\inf_{H(I)} \text{pr}\left\{p_{(1)} > c_{mm}\alpha, \ldots, p_{(m)} > c_{m1}\alpha \mid H(I)\right\} \geqslant 1 - \alpha, \quad (1)$$

where the infimum is taken over all nonempty intersection hypotheses $H(I) = \bigcap_{i \in I} H_i$ with $I \subseteq \{1, \ldots, n\}$. Then the α-level local test rejects $H(I)$ if

$$p_{(m-i+1)} \leqslant c_{mi}\alpha \text{ for at least one } i = 1, \ldots, m. \quad (2)$$

The critical constants c_{mi} can be determined by using the fact that under $H(I)$, $p_{(1)} \leqslant \cdots \leqslant p_{(m)}$ are order statistics from a uniform [0, 1] distribution.

In general, a closure procedure based on the local test (2) does not have a simple stepwise shortcut. Liu (1996) showed that if C denotes an $n \times n$ lower-triangular matrix with entries c_{mi} $(i = 1, \ldots, m; m = 1, \ldots, n)$, then the closure procedure has a step-down shortcut if the row entries of C are equal and a step-up shortcut if the column entries are equal. In the former case, the critical constants for comparing $p_{(n)}, \ldots, p_{(1)}$ are α times the first column, $(c_{11}, \ldots, c_{n1})$, of C and in the latter case, they are α times the last row, $(c_{n1}, \ldots, c_{nn})$, of C.

The Holm procedure is a step-down shortcut to the closed procedure that uses the Bonferroni test as a local test for every intersection hypothesis. The Bonferroni test rejects $H(I)$ if $\min_{i \in I} p_i \leqslant \alpha/m$ where $m = |I|$. The corresponding C-matrix has equal row entries, $c_{mi} = c_m = 1/m$ $(m = 1, \ldots, n)$, and hence the closed procedure has a step-down shortcut that is the Holm procedure.

The Hommel procedure is a shortcut to the closed procedure that uses the Simes (1986) test as a local test of every intersection hypothesis. The Simes test rejects $H(I)$ if at least one $p_{(i)} \leqslant i\alpha/m$ $(i = 1, \ldots, m)$. It is based on the identity

$$\text{pr}\left\{ p_{(i)} > \frac{i\alpha}{m}(i = 1, \ldots, m) \middle| H(I) \right\} = 1 - \alpha$$

and hence is an exact α-level test.

Table 1. *Critical constants for the Rom procedure*

α	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}
1%	1·000	0·500	0·334	0·251	0·201	0·167	0·144	0·126	0·112	0·100
5%	1·000	0·500	0·338	0·254	0·204	0·170	0·146	0·128	0·114	0·102

The entries of the C-matrix for the Hommel procedure are $c_{mi} = (m - i + 1)/m$ for $i = 1, \ldots, m$. These entries are not constant across either rows or columns, so the Hommel procedure does not have a simple step-down or step-up structure. It operates as follows: at Step 1, if $p_{(n)} \leqslant \alpha$, then reject all hypotheses; otherwise accept $H_{(n)}$ and go to Step 2. In general, at Step i, if at least one $p_{(j)} \leqslant (i + j - n)\alpha/i$ for $j = n, \ldots, n - i + 1$, then reject any hypothesis with a p-value $\leqslant \alpha/(i - 1)$; otherwise accept $H_{(n-i+1)}$ and go to Step $i + 1$. At Step n, reject $H_{(1)}$ if $p_{(1)} \leqslant \alpha/(n - 1)$. Thus, if the procedure goes past Step 1 by accepting $H_{(n)}$, then at Step 2 it tests whether $p_{(n-1)} \leqslant \alpha/2$ and if so it rejects any hypothesis with $p_i \leqslant \alpha$, i.e., all the remaining hypotheses, and stops testing. Otherwise at the next step it tests whether $p_{(n-1)} \leqslant 2\alpha/3$ or $p_{(n-2)} \leqslant \alpha/3$. If either of these inequalities holds, then it rejects any hypothesis with $p_i \leqslant \alpha/2$ and so on.

The Hochberg procedure is a conservative shortcut to the Hommel procedure. However, it is more easily derived using the local test with a C-matrix having constant column entries $c_{mi} = c_i = 1/i$ $(i = 1, \ldots, n)$. Therefore the Hochberg procedure has a simple step-up structure. The entries in this matrix are more conservative than those in the Hommel procedure C-matrix since $1/i \leqslant (m - i + 1)/m$ for $i = 1, \ldots, m$ with equalities only for $i = 1$ and $i = m$, thus resulting in a strict inequality in (1). So the Hochberg procedure is uniformly less powerful than the Hommel procedure. On the other hand, it is uniformly more powerful than the Holm procedure because $1/i \geqslant 1/m$ for $i = 1, \ldots, m$, with equality holding only for $i = m$.

The Rom procedure is identical to the Hochberg procedure except that it uses the exact and hence sharper critical constants. At the ith step, it compares $p_{(n-i+1)}$ with the exact critical constant $c_i\alpha$ instead of the conservative critical constant α/i used by the Hochberg procedure, where the c_is satisfy the equation

$$\text{pr}\left\{p_{(1)} > c_n\alpha, \ldots, p_{(n)} > c_1\alpha \mid H(I)\right\} = 1 - \alpha.$$

The c_is can be evaluated iteratively as follows: we have $c_1 = 1$ since under H_1, $\text{pr}\,(p_1 > \alpha) = 1 - \alpha$. Next, for $n = 2$, we can obtain c_2 by solving the equation

$$\text{pr}\left(p_{(2)} > \alpha, p_{(1)} > c_2\alpha\right) = 1 - \alpha,$$

which gives $c_2 = 1/2$. In general, the equation for obtaining c_i is

$$c_i = \frac{1}{i}\left\{1 + \sum_{j=1}^{i-2}\alpha^j - \sum_{j=1}^{i-2}\binom{i}{j}c_{j+1}^{i-j}\alpha^{i-j-1}\right\} \qquad (i = 2, \ldots, n). \tag{3}$$

Recursive computations give

$$c_1 = 1, \quad c_2 = \frac{1}{2}, \quad c_3 = \frac{1 + \alpha/4}{3}, \quad c_4 = \frac{1 + \alpha/3 + \alpha^2/6 - \alpha^3/24}{4}.$$

These c_is are functions of α but not of n, i.e., for any given n, the same c_is are used for all $i = 1, \ldots, n$. Values of the c_i $(i = 1, \ldots, 10)$ are given in Table 1.

3. A CLASS OF HYBRID PROCEDURES

3·1. *Hybrid*-0 *procedure*

This special case of the hybrid procedure operates as follows: at Step 1, if $p_{(n)} \leqslant \alpha$, then reject all hypotheses; otherwise accept $H_{(n)}$ and go to the next step. In general, at Step i, if $p_{(n-i+1)} \leqslant c_i\alpha$, where $c_i = (i+1)/(2i)$, then reject any hypothesis with a p-value $\leqslant d_i\alpha$ where $d_i = 1/i$ and stop; otherwise accept $H_{(n-i+1)}$ and continue to the next step ($i = 1, \ldots, n-1$). At Step n, reject $H_{(1)}$ if $p_{(1)} \leqslant \alpha/n$; otherwise accept $H_{(1)}$ and stop. In Theorem 4 we show that the critical constants of the hybrid-0 procedure are the zeroth-order approximations to the exact critical constants (c_i, d_i) of the general hybrid procedure defined in § 3·2. Furthermore, in Theorem 5 we show that hybrid-0 controls the familywise error rate.

The entries of the C-matrix for the hybrid-0 procedure are $c_{mi} = c_i = (i+1)/(2i)$ for $i = 1, \ldots, m-1$ and $c_{mm} = d_m = 1/m$ for $m = 1, \ldots, n$. It is readily seen that for $n = 2$, the hybrid-0, Hochberg, Hommel and Rom procedures are identical and they all control the familywise error rate exactly at level α. The hybrid-0 procedure is uniformly more powerful than the Hochberg procedure for $n \geqslant 3$ because its critical constants c_{mi} are greater than the critical constants $c_{mi} = 1/i$ of the Hochberg procedure for $i = 1, \ldots, m$ with equalities only for $i = 1$ and $i = m$. A more detailed check shows that the hybrid-0 procedure is uniformly more powerful than the Hommel procedure for $n = 3$ and 4, but not necessarily for $n \geqslant 5$, as Example 3 below shows.

Example 1. Suppose $n = 3$, $\alpha = 0{\cdot}05$, and $p_{(1)} = 0{\cdot}02$, $p_{(2)} = 0{\cdot}035$, $p_{(3)} = 0{\cdot}06$. The Simes test accepts the overall intersection hypothesis $H_0 = H_{(1)} \bigcap H_{(2)} \bigcap H_{(3)}$ since $p_{(1)} > \alpha/3$, $p_{(2)} > 2\alpha/3$, $p_{(3)} > \alpha$. Therefore the Hommel procedure does not reject any hypothesis. However, the hybrid-0 procedure rejects $H_{(1)}$ at Step 2 since $p_{(2)} \leqslant 3\alpha/4 = 0{\cdot}0375$ and $p_{(1)} \leqslant \alpha/2 = 0{\cdot}025$.

Example 2. Suppose $n = 4$, $\alpha = 0{\cdot}05$ and $p_{(1)} = 0{\cdot}02$, $p_{(2)} = 0{\cdot}03$, $p_{(3)} = 0{\cdot}035$, $p_{(4)} = 0{\cdot}06$. Applying the Hommel procedure as a closed procedure, we see that the Simes test rejects H_0. Next it rejects two out of four three-way intersections, namely, $H_{(1)} \bigcap H_{(2)} \bigcap H_{(3)}$ and $H_{(1)} \bigcap H_{(2)} \bigcap H_{(4)}$. So the Hommel procedure does not reject any of the elementary hypotheses since they are all included in one of the two unrejected intersection hypotheses. But the hybrid-0 procedure rejects $H_{(1)}$ at Step 2, since $p_{(3)} \leqslant 3\alpha/4 = 0{\cdot}0375$ and $p_{(1)} \leqslant \alpha/2 = 0{\cdot}025$.

Example 3. Suppose $n = 5$, $\alpha = 0{\cdot}05$, and $p_{(1)} = 0{\cdot}011$, $p_{(2)} = 0{\cdot}032$, $p_{(3)} = 0{\cdot}034$, $p_{(4)} = 0{\cdot}039$, $p_{(5)} = 0{\cdot}06$. It is easy to check that the hybrid-0 procedure does not reject any hypotheses since

$$p_{(5)} > 0{\cdot}05, \quad p_{(4)} > \frac{3}{4}(0{\cdot}05), \quad p_{(3)} > \frac{2}{3}(0{\cdot}05), \quad p_{(2)} > \frac{5}{8}(0{\cdot}05), \quad p_{(1)} > \frac{1}{5}(0{\cdot}05).$$

But the Hommel procedure rejects $H_{(1)}$ since at the final step, $p_{(4)} \leqslant (4/5)(0{\cdot}05) = 0{\cdot}04$ and $p_{(1)} \leqslant 0{\cdot}05/4 = 0{\cdot}0125$.

3·2. *General hybrid procedure*

Fix two monotone sequences of critical constants $c_1 \geqslant \cdots \geqslant c_n$ and $d_1 \geqslant \cdots \geqslant d_n$ such that $1 \geqslant c_i \geqslant d_i$ for all i. A general hybrid procedure operates as follows.

Step i ($i = 1, \ldots, n-1$). If $p_{(n-i+1)} \leqslant c_i\alpha$, then reject any hypothesis with a p-value $\leqslant d_i\alpha$ and stop; otherwise accept $H_{(n-i+1)}$ and go to Step $i+1$.

Step n. If $p_{(1)} \leqslant d_n\alpha$, then reject $H_{(1)}$; otherwise accept $H_{(1)}$ and stop.

From the last step, it follows that without loss of generality, we can take $c_n = d_n$.

Remark 1. The Hochberg and the Rom procedures are special cases of the general hybrid procedure. For the Hochberg procedure we have $c_i = d_i = 1/i$ while for the Rom procedure the $c_i = d_i$ are determined from equation (3). The Hommel procedure uses different critical constants $c_{ij} = (i + j - n)/j$ at each step i for testing $p_{(j)}$ $(j = n, n-1, \ldots, n-i+1)$; furthermore, it uses $d_1 = 1$ and $d_i = 1/(i-1)$ for $i = 2, \ldots, n$.

3·3. *Properties of the hybrid procedure*

We now state two theorems that show the closure and the consonance properties of the hybrid procedure.

THEOREM 1. *The hybrid procedure is a shortcut to the closed procedure that rejects any nonempty intersection hypothesis* $H(I) = \bigcap_{i \in I} H_i$ *where* $I = \{1, \ldots, m\}$ *and* $m \leqslant n$ *if one of the following mutually exclusive events occurs:*

$$E_i(m) = \begin{cases} p_{(m)} \leqslant \alpha & (i = 1), \\ p_{(m)} > \alpha,\ p_{(m-1)} > c_2\alpha, \ldots, p_{(m-i+2)} > c_{i-1}\alpha, & \\ p_{(m-i+1)} \leqslant c_i\alpha,\ p_{(1)} \leqslant d_i\alpha & (i = 2, \ldots, m). \end{cases} \tag{4}$$

Since the hybrid procedure is a closed procedure with its critical constants $\{(c_i, d_i),\ i = 1, \ldots, n\}$ chosen according to Theorem 3 so that the local test of every intersection hypothesis $H(I)$ is of level α, it follows that the hybrid procedure strongly controls the familywise error rate at level α.

THEOREM 2. *The hybrid procedure, viewed as a closed procedure, satisfies the consonance condition of Hommel et al. (2007).*

3·4. *Critical constants of the hybrid procedure*

An alternative representation of the event $\bigcup_{i=1}^{m} E_i(m)$ that is useful for evaluating the critical constants is $\bigcup_{i=1}^{m} E_i^*(m)$ where the $E_i^*(m)$ are mutually exclusive events defined as follows:

$$E_i^*(m) = \begin{cases} p_{(m)} \leqslant \alpha,\ p_{(1)} > d_m\alpha & (i = 1), \\ p_{(m)} > \alpha,\ p_{(m-1)} > c_2\alpha, \ldots, p_{(m-i+2)} > c_{i-1}\alpha, & \\ p_{(m-i+1)} \leqslant c_i\alpha,\ d_m\alpha < p_{(1)} \leqslant d_i\alpha & (i = 2, \ldots, m-1), \\ p_{(1)} \leqslant d_m\alpha & (i = m). \end{cases} \tag{5}$$

THEOREM 3. *Let* $A_i^*(m) = \text{pr}\{E_i^*(m) \mid H(I)\}$. *Then the exact critical constants* (c_i, d_i) *can be determined from the equation*

$$\text{pr}\,\{\text{Reject } H(I) \mid H(I)\} = \sum_{i=1}^{m} A_i^*(m) = \alpha$$

for $m = 1, \ldots, n$ *where*

$$A_i^*(m) = \begin{cases} \alpha^m (1 - d_m)^m & (i = 1), \\ B_i(m) \sum_{k=1}^{m-i+1} \binom{m-i+1}{k} (c_i - d_i)^{m-i+1-k} (d_i - d_m)^k \alpha^{m-i+1}, & (i = 2, \ldots, m-1), \\ 1 - (1 - d_m\alpha)^m & (i = m). \end{cases} \tag{6}$$

Table 2. *Critical constants for the hybrid procedure*

c_i	Zeroth-order approx.	1st-Order approx.		Exact value	
		$\alpha = 5\%$	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 1\%$
c_1	1	1	1	1	1
c_2	0·75	0·75	0·75	0·75	0·75
c_3	0·667	0·670	0·667	0·670	0·667
c_4	0·625	0·629	0·626	0·629	0·626
c_5	0·600	0·604	0·601	0·604	0·601
c_6	0·583	0·587	0·584	0·587	0·584
c_7	0·571	0·576	0·572	0·576	0·572
c_8	0·563	0·567	0·563	0·567	0·563
c_9	0·556	0·560	0·556	0·560	0·556
c_{10}	0·550	0·554	0·551	0·554	0·551

Here

$$B_i(m) = \sum \binom{m}{j_1, \ldots, j_i} (1-\alpha)^{j_1} (c_1 - c_2)^{j_2} \cdots (c_{i-2} - c_{i-1})^{j_{i-1}} \alpha^{i-1-j_1}, \tag{7}$$

where

$$\binom{m}{j_1, \ldots, j_i} = \frac{m!}{j_1! \cdots j_i!}$$

and the sum is over all $j_1, \ldots, j_i$ *such that* $j_1 + \cdots + j_{i-1} = i-1$, $j_i = m - i + 1$, $j_1 \geqslant 1$ *and* $j_k \geqslant 0$ *for* $k > 1$.

If $m = n = 1$, then we require pr $(p_1 \leqslant d_1\alpha) = d_1\alpha = \alpha$, so $c_1 = d_1 = 1$. For $n \geqslant 2$, these equations must be recursively solved. Two issues arise in solving these equations. First, they cannot be simultaneously solved for both c_i and d_i unless we impose some restrictions. Theorem 4 gives the rationale for the restriction $d_i = 1/i$. Second, even after imposing this restriction it appears difficult to show analytically that the solutions c_i satisfy the monotonicity condition $c_1 \geqslant \cdots \geqslant c_n$. In general, this is a difficult problem even for the simple Rom step-up procedure and was solved in a completely different context by Dalal & Mallows (1988). However, the numerical calculations shown in Table 2 show that the exact c_is are in fact monotone. For the hybrid-0 procedure, the critical constants $c_i = (i+1)/(2i)$ are obviously monotone.

THEOREM 4. *Asymptotically, as* $\alpha \to 0$,

$$c_i = \frac{d_i^2}{2(d_i - d_{i+1})} + \frac{1}{\alpha}\left\{\frac{1/(i+1) - d_{i+1}}{i(d_i - d_{i+1})}\right\} + o(1), \tag{8}$$

$$d_i = \frac{1}{i} + \alpha\left\{\frac{1}{2(i-1)} - \frac{c_{i-1}}{i}\right\} + o(\alpha). \tag{9}$$

Therefore, in order that the c_i *remain finite as* $\alpha \to 0$, *the second term in* (8) *must be zero and hence* $d_{i+1} = 1/(i+1)$, *which gives*

$$c_i = \frac{i+1}{2i} + o(1), \quad d_i = \frac{1}{i} + o(\alpha). \tag{10}$$

We will refer to the critical constants given by (10) as the zeroth-order approximations. The exact c_is can be calculated iteratively by setting $c_1 = 1$, $c_n = 1/n$, $d_i = 1/i$ $(i = 1, \ldots, n)$ and then given c_1 calculate c_2 or in general given $c_1, c_2, \ldots, c_{i-1}$ for $i \leqslant n-2$ calculate

$$c_i = \frac{2i+1}{2i(i+1)} + \frac{\alpha - \left[\left\{i\alpha/(i+1)\right\}^{i+1} + 1 - \left\{1-\alpha/(i+1)\right\}^{i+1} + \sum_{j=2}^{i-1} A_j^*(i+1)\right]}{\{2\alpha^2/i(i+1)\}B_i(i+1)}, \tag{11}$$

where the $A_j^*(i+1)$ and $B_i(i+1)$ are defined in (6) and (7), respectively; note that both require prior calculation of $c_1, \ldots, c_{i-1}$. The first-order approximation to the above formula is

$$c_i = \frac{i+1}{2i} + \frac{\alpha}{12}\left\{1 - \frac{1}{(i-1)^2}\right\} + o(\alpha). \tag{12}$$

We denote the hybrid procedure that uses the exact critical constants (11) by hybrid-∞ and the hybrid procedure that uses the first-order approximation (12) by hybrid-1.

The first five first-order approximations are

$$c_1 = 1, \quad c_2 = \frac{3}{4}, \quad c_3 = \frac{2}{3} + \frac{\alpha}{16}, \quad c_4 = \frac{5}{8} + \frac{2\alpha}{27}, \quad c_5 = \frac{3}{5} + \frac{5\alpha}{64}.$$

From Table 2 we see that they are equal to the exact critical constants up to the third decimal place. Even though the zeroth-order approximations differ in the third decimal place, the simulated familywise error rates of hybrid-∞ and hybrid-0 are quite close. Furthermore, the zeroth-order approximations are easier to use because they do not depend upon α and need no tabling, while the exact and the first-order approximations do. Therefore we will focus on the hybrid-0 procedure henceforth.

4. PROPERTIES OF THE HYBRID-0 PROCEDURE

It is clear that hybrid-∞ controls the familywise error rate exactly. However, it is not clear that hybrid-0 controls the familywise error rate conservatively. The calculations in Table 2 indicate that the zeroth-order approximations are always less than the exact values, so we may conjecture that hybrid-0 must be conservative. The next theorem proves this result analytically.

THEOREM 5. *The hybrid*-0 *procedure strongly controls the familywise error rate at level α, the control being exact for $n = 2$ and* 3 *and conservative for $n \geqslant 4$.*

The p-values adjusted for multiplicity are often helpful since they can be directly compared with any α to make rejection decisions. The adjusted p-values for the hybrid-0 procedure equal

$$\tilde{p}_{[i]} = \min_{j=1,\ldots,i} \left\{\max\left(\frac{p_{[j]}}{c_j}, \frac{p_{[i]}}{d_j}\right)\right\} \ (i = 1, \ldots, n), \tag{13}$$

where $p_{[1]} \geqslant \cdots \geqslant p_{[n]}$ denote the p-values in decreasing order and $H_{[1]}, \ldots, H_{[n]}$ denote the corresponding hypotheses.

Remark 2. The derivation of (13) requires that the critical constants (c_j, d_j) do not depend upon α, which is true for the hybrid-0 procedure, but not for the hybrid-1 and hybrid-∞ procedures. Numerical search techniques must be used for these two procedures.

Table 3. *Exact familywise error rate expressions under the overall null hypothesis and independence for Hommel and Hochberg procedures*

n	Procedure	Familywise error rate
3	Hommel	$\alpha - \frac{1}{12}\alpha^2(1-\alpha)$
	Hochberg	$\alpha - \frac{1}{4}\alpha^2(1-\alpha)$
4	Hommel	$\alpha - \frac{1}{6}\alpha^2(1-\alpha)^2 - \frac{35}{144}\alpha^3(1-\alpha)$
	Hochberg	$\alpha - \frac{1}{3}\alpha^2(1-\alpha)^2 - \frac{7}{12}\alpha^3(1-\alpha)$
5	Hommel	$\alpha - \frac{9}{40}\alpha^2(1-\alpha)^3 - \frac{6359}{10800}\alpha^3(1-\alpha)^2 - \frac{157219}{259200}\alpha^4(1-\alpha)$
	Hochberg	$\alpha - \frac{3}{8}\alpha^2(1-\alpha)^3 - \frac{113}{108}\alpha^3(1-\alpha)^2 - \frac{2053}{2160}\alpha^4(1-\alpha)$
n	Hommel	$\alpha - \frac{(n-2)^2}{2(n-1)n}\alpha^2 + o(\alpha^2)$
	Hochberg	$\alpha - \frac{n-2}{2(n-1)}\alpha^2 + o(\alpha^2)$

5. Familywise error rates

All results given in this section are under the overall null hypothesis H_0. The following exact expressions for the familywise error rate of the hybrid-0 procedure under independence for $n = 3$, 4, 5 can be derived from Theorem 3:

$$\text{Familywise error rate} = \begin{cases} \alpha & (n=3), \\ \alpha - \frac{1}{16}\alpha^3(1-\alpha) & (n=4), \\ \alpha - \frac{2}{27}\alpha^3(1-\alpha)^2 - \frac{83}{540}\alpha^4(1-\alpha) & (n=5). \end{cases} \quad (14)$$

These expressions show that the familywise error rate of hybrid-0 is strictly less than α for $n = 4$ and 5. Similar expressions for the Hommel and Hochberg procedures are given without derivations in Table 3. These expressions were used to calculate the achieved familywise error rates of these procedures reported in Table 4 for $\rho = 0$.

We conducted a large simulation study to evaluate the familywise error rates of the competing procedures for correlated test statistics. The correlated p-values were generated by first generating n standard normal random variables $z_1, \ldots, z_n$ with common correlation ρ and setting $p_i = 1 - \Phi(z_i)$ where $\Phi(\cdot)$ denotes the standard normal cumulative distribution function. A total of 10^9 random samples of $z_1, \ldots, z_n$ were generated for each simulation run and the four procedures were applied to the same data. The standard error of each simulation estimate for nominal $\alpha = 5\%$ is less than 0·0007%. The results for the independence case and for $\rho = 0{\cdot}5$ and 0·7 are presented in Table 4.

Next we turn to the negative dependence case. The common correlation ρ must be greater than $-1/(n-1)$ to assure that the distribution of $z_1, \ldots, z_n$ is not degenerate. We used the same simulation set-up as for the positive dependence case. So if the estimate exceeds 5·0014% then we can conclude that the familywise error rate is not controlled at the nominal 5% level. The results for some selected cases are given in Table 5.

Table 4. *Familywise error rates* (%) *of Hochberg, Rom, Hommel, and hybrid*-0 *procedures under independence and common positive correlation* ($\alpha = 5\%$)

ρ	n	Procedures			
		HC	RM	HM	HH-0
0	3	4·940	5·000	4·980	5·000
	5	4·907	5·000	4·945	4·999
	7	4·897	5·000	4·927	4·999
0·5	3	4·424	4·470	4·539	4·593
	5	4·048	4·118	4·208	4·370
	7	3·839	3·913	3·987	4·218
0·7	3	3·988	4·024	4·134	4·203
	5	3·374	3·429	3·601	3·798
	7	3·036	3·092	3·260	3·539

HC, Hochberg procedure; RM, Rom procedure; HM, Hommel procedure; HH-0, hybrid-0 procedure.

Table 5. *Familywise error rates* (%) *of Hochberg, Rom, Hommel, and hybrid*-0 *procedures under common negative correlation* ($\alpha = 5\%$)

n	ρ	Procedures			
		HC	RM	HM	HH-0
3	−0·4	5·000	5·062	5·003	5·005
5	−0·2	4·985	5·081	4·996	5·015
7	−0·15	4·973	5·079	4·985	5·016
10	−0·1	4·956	5·068	4·967	5·016

HC, Hochberg procedure; RM, Rom procedure; HM, Hommel procedure; HH-0, hybrid-0 procedure.

These simulations suggest the following conclusions. All four procedures become more conservative as ρ increases and/or n increases. For $\rho \geqslant 0$, the hybrid-0 procedure achieves the familywise error rate that is closest to the nominal $\alpha = 5\%$. Therefore it is the least conservative among all four procedures. For $\rho < 0$, the Rom and the hybrid-0 procedures are anti-conservative with the former being more so; nonetheless, the nominal α is exceeded by less than 0·1% for the Rom procedure and by less than 0·02% for the hybrid-0 procedure. The Hommel procedure does not control the error rate only for $n = 3$, $\rho = -0{\cdot}4$, while the Hochberg procedure controls the error rate in all the cases studied. The reason is that the Hochberg procedure is inherently conservative.

6. POWER

We used two definitions of power: familywise power or F-power, which is the probability of rejecting at least one false hypothesis, and average power or A-power, which is the expected proportion of rejected false hypotheses. Denote the number of false hypotheses by $m = 1, \ldots, n$. For $m = 1$, the two definitions coincide. The Hochberg procedure is known to be the least powerful of the four procedures whose familywise error rates were studied in § 5, so we use it as a benchmark for the other three procedures. The details of the simulation study are as follows. We generated the p_i-values from the normally distributed z_is as described in the previous section with $z_i \sim N(0, 1)$ for the true hypotheses and $z_i \sim N(\delta, 1)$ for the false hypotheses with $\delta = 2{\cdot}0$. Here we report the results for $n = 5$, 10 and 20, but we also studied $n = 50$ and 100, the results for which are not reported here to save space. For each n, we considered $\rho = 0$, 0·5 and 0·7, and

Table 6. *Simulated F-powers* (%) *for Hochberg, Rom, Hommel, and hybrid*-0 *procedures* ($\alpha =$ 5%)

		$\rho = 0$				$\rho = 0{\cdot}5$				$\rho = 0{\cdot}7$			
n	m	HC	RM	HM	HH-0	HC	RM	HM	HH-0	HC	RM	HM	HH-0
5	1	37·4	37·7*	37·5	37·6	37·3	37·6*	37·4	37·5	37·3	37·5*	37·3	37·4
	3	75·6	75·9	76·6	77·5*	62·1	62·4	63·2	64·1*	56·1	56·4	57·3	58·3*
	5	90·7	90·9	91·9	92·6*	72·6	72·8	74·0	75·0*	64·4	64·6	66·1	67·2*
10	2	48·6	49·0	48·9	49·5*	42·2	42·5	42·5	43·1*	38·9	39·2	39·2	39·8*
	6	86·4	86·7	87·4	88·8*	64·7	65·0	65·7	67·4*	55·2	55·5	56·4	58·5*
	10	96·4	96·6	97·2	98·0*	73·4	73·7	74·9	76·8*	61·9	62·2	64·0	66·4*
20	4	61·1	61·5	61·4	62·6*	46·0	46·4	46·3	47·3*	39·5	39·8	39·8	40·8*
	12	94·1	94·3	94·7	95·9*	66·2	66·6	67·0	69·2*	53·7	54·1	54·8	57·6*
	20	99·1	99·2	99·4	99·7*	73·9	74·2	75·0	77·6*	59·6	59·9	61·5	64·9*

m, the number of false hypotheses; HC, Hochberg procedure; RM, Rom procedure; HM, Hommel procedure; HH-0, hybrid-0 procedure. An asterisk indicates the highest power among the four procedures for a given combination of m, n and ρ.

Table 7. *Simulated A-powers* (%) *for Hochberg, Rom, Hommel, and hybrid*-0 *procedures* ($\alpha =$ 5%)

		$\rho = 0$				$\rho = 0{\cdot}5$				$\rho = 0{\cdot}7$			
n	m	HC	RM	HM	HH-0	HC	RM	HM	HH-0	HC	RM	HM	HH-0
5	1	37·4	37·7*	37·5	37·6	37·3	37·6*	37·4	37·5	37·3	37·5*	37·3	37·4
	3	40·4	40·7	41·1	41·6*	40·8	41·1	41·5	42·1*	40·8	41·1	41·7	42·3*
	5	47·4	47·5	48·4	48·9*	48·9	49·0	49·8	50·3*	49·7	49·8	50·6	51·1*
10	2	28·7	28·9	28·8	29·2*	28·8	29·1	29·0	29·4*	28·8	29·1	29·0	29·4*
	6	30·5	30·8	31·3	32·5*	31·4	31·7	32·4	33·7*	31·7	31·9	32·9	34·3*
	10	34·2	34·5	36·5	38·4*	38·8	39·1	41·0	42·5*	40·8	41·0	43·1	44·5*
20	4	21·3	21·6	21·5	22·0*	21·5	21·7	21·7	22·4*	21·6	21·8	21·8	22·5*
	12	22·3	22·6	23·0	24·6*	23·3	23·5	24·4	26·3*	23·7	23·9	25·0	27·1*
	20	23·7	24·0	25·8	28·9*	28·9	29·2	32·4	34·9*	32·0	32·2	35·8	38·0*

m, the number of false hypotheses; HC, Hochberg procedure; RM, Rom procedure; HM, Hommel procedure; HH-0, hybrid-0 procedure. An asterisk indicates the highest power among the four procedures for a given combination of m, n and ρ.

$m/n = 0{\cdot}2$, $0{\cdot}6$, $1{\cdot}0$. The rest of the simulation set-up was the same as was used for familywise error rates. The results for F-power are given in Table 6 and for A-power in Table 7.

These simulations suggest the following conclusions. The hybrid-0 procedure has the highest F-power as well as the highest A-power in all the cases studied except for $n = 5$, $m = 1$, where the Rom procedure has a marginally higher power. The F-power and A-power gains of the hybrid-0 procedure over the Hommel procedure, which is generally the second best procedure, are of the same order of magnitude. For $m = n = 20$, the highest F-power gain is 3·4% when $\rho = 0{\cdot}7$ while the highest A-power gain is 3·1% when $\rho = 0$. Compared to the Hochberg procedure, which is the most commonly used procedure in practice, the A-power gains of the Hybrid-0 procedure are nearly twice as high reaching 5% to 6% when $m = n = 20$. These gains increase with n reaching 8% to 9% when $m = n = 50$ and 100. In general, the Rom procedure is the least powerful of the three competing procedures having the least power gain over the Hochberg procedure in almost all cases.

Table 8. *Adjusted p-values*

i	p_i	BF	BH	HC	HM	HH-0
1	0·002	0·020*	0·020*	0·020*	0·018*	0·014*
2	0·005	0·050	0·045*	0·045*	0·032*	0·030*
3	0·007	0·070	0·056	0·049*	0·042*	0·037*
4	0·007	0·070	0·056	0·049*	0·042*	0·037*
5	0·009	0·090	0·056	0·054	0·045*	0·038*
6	0·022	0·220	0·110	0·060	0·054	0·048*
7	0·024	0·240	0·110	0·060	0·054	0·048*
8	0·035	0·350	0·110	0·060	0·060	0·060
9	0·036	0·360	0·110	0·060	0·060	0·060
10	0·060	0·600	0·110	0·060	0·060	0·060

An asterisk indicates that the adjusted p-value is significant at $\alpha = 5\%$.

7. Example

This numerical example illustrates the differences among the Hochberg, Hommel, and hybrid-0 procedures. We have also included the Bonferroni and Holm procedures although they are not competitive with the others. The raw p-values for $n = 10$ hypotheses and the adjusted p-values $\tilde{p}_i$ computed using the following formulae are shown in Table 8. Here all $\tilde{p}_i$ are truncated at 1.

The formulae are as follows: Bonferroni procedure, $\tilde{p}_i = np_i$; Holm procedure, $\tilde{p}_{(i)} = \max\{(n - i + 1)p_{(i)}, \tilde{p}_{(i-1)}\}$; Hochberg procedure, $\tilde{p}_{(i)} = \min\{(n - i + 1)p_{(i)}, \tilde{p}_{(i+1)}\}$; Hommel procedure, algorithm in Wright (1992); and hybrid-0 procedure, formula (13).

The adjusted p-values that are significant at $\alpha = 5\%$ are indicated by asterisks in the table. We get one significance for the Bonferroni procedure, two for the Holm procedure, four for the Hochberg procedure, five for the Hommel procedure and seven for the hybrid-0 procedure. This illustrates the power advantage of the hybrid-0 procedure.

Acknowledgement

We thank Dr Willi Maurer for bringing Romano et al. (2011) to our attention. We also thank two referees, an associate editor and the editor for helpful comments and corrections. Support from Prosoft, Inc. is gratefully acknowledged.

Supplementary material

Supplementary material available at *Biometrika* online includes proofs of the theorems and derivations of the formulae in this paper.

References

Dalal, S. R. & Mallows, C. L. (1988). When should one stop testing software? *J. Am. Statist. Assoc.* **83**, 872–9.

Gabriel, K. R. (1969). Simultaneous test procedures – Some theory of multiple comparisons. *Ann. Math. Statist.* **40**, 224–50.

Hochberg, Y. (1988). A sharper Bonferroni procedure for multiple tests of significance. *Biometrika* **75**, 800–2.

Hochberg, Y. & Tamhane, A. C. (1987). *Multiple Comparison Procedures*. New York: John Wiley.

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

Hommel, G. (1988). A stagewise rejective multiple test procedure based on a modified Bonferroni test. *Biometrika* **75**, 383–6.

Hommel, G., Bretz, F. & Maurer, W. (2007). Powerful short-cuts for multiple testing procedures with special reference to gatekeeping strategies. *Statist. Med.* **26**, 4063–73.

LIU, W. (1996). On multiple tests of a non-hierarchical finite family of hypotheses. *J. R. Statist. Soc.* B **58**, 455–61.
MARCUS, R., PERITZ, E. & GABRIEL, K. R. (1976). On closed testing procedures with special reference to ordered analysis of variance. *Biometrika* **63**, 655–60.
ROM, D. M. (1990). A sequentially rejective test procedure based on a modified Bonferroni inequality. *Biometrika* **77**, 663–5.
ROM, D. M. (2013). An improved Hochberg procedure for multiple tests of significance. *Br. J. Math. Statist. Psychol.* **66**, 189–96.
ROMANO, J. P., SHAIKH, A. M. & WOLF, M. (2011). Consonance and the closure method in multiple testing. *Int. J. Biostatist.* **7**, 1–25.
SARKAR, S. K. (1998). Some probability inequalities for ordered MTP2 random variables: a proof of the Simes conjecture. *Ann. Statist.* **26**, 494–504.
SARKAR, S. K. & CHANG, C. K. (1997). The Simes method for multiple hypothesis testing with positively dependent test statistics. *J. Am. Statist. Assoc.* **92**, 1601–8.
SIMES, R. J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika* **73**, 751–4.
WESTFALL, P. H., TOBIAS, R. D., ROM, D., WOLFINGER, R. D. & HOCHBERG, Y. (1999). *Multiple Comparisons and Multiple Tests Using the SAS System*. Cary, North Carolina, USA: SAS Institute.
WRIGHT, S. P. (1992). Adjusted *p*-values for simultaneous inference. *Biometrics* **48**, 1005–13.

[*Received July* 2013. *Revised April* 2014]