Biometrika (2008), **95**, 2, *pp.* 279–294 doi: 10.1093/biomet/asn018

Printed in Great Britain

On weighted Hochberg procedures

BY AJIT C. TAMHANE

Department of Industrial Engineering and Management Sciences, Northwestern University, 2145 Sheridan Road, Evanston, Illinois 60208, U.S.A.

ajit@iems.northwestern.edu

AND LINGYUN LIU

Department of Statistics, Northwestern University, 2008 Sheridan Road, Evanston, Illinois 60208, U.S.A.

lingyunliu@northwestern.edu

SUMMARY

We consider different ways of constructing weighted Hochberg-type step-up multiple test procedures including closed procedures based on weighted Simes tests and their conservative step-up short-cuts, and step-up counterparts of two weighted Holm procedures. It is shown that the step-up counterparts have some serious pitfalls such as lack of familywise error rate control and lack of monotonicity in rejection decisions in terms of p-values. Therefore an exact closed procedure appears to be the best alternative, its only drawback being lack of simple stepwise structure. A conservative step-up short-cut to the closed procedure may be used instead, but with accompanying loss of power. Simulations are used to study the familywise error rate and power properties of the competing procedures for independent and correlated p-values. Although many of the results of this paper are negative, they are useful in highlighting the need for caution when procedures with similar pitfalls may be used.

Some key words: Bonferroni test; Closed procedure; Familywise error rate; Holm procedure; Multiple comparisons; p-Value; Simes test; Step-down procedure; Step-up procedure.

1. INTRODUCTION

Hochberg's (1988) procedure is widely used in practice because of its simple step-up testing algorithm and higher power in the class of p-value-based multiple test procedures, compared to the Holm (1979) step-down procedure, for example. Weighted versions of the Bonferroni and Holm procedures, as well as that of a sharper Bonferroni-type procedure due to Simes (1986), have been proposed for multiple testing problems; see Holm (1979), Hochberg & Liberman (1994) and Benjamini & Hochberg (1997). However, a weighted version of the Hochberg procedure has not been proposed, although such a procedure would potentially inherit similar power advantages over its competitors. The purpose of the present paper is to study alternative ways of constructing weighted Hochberg-type procedures and make recommendations for their use. The study helps us to highlight the pitfalls associated with some of the weighted step-up procedures.

Hypotheses may be differentially weighted because of their different importance, and to obtain higher power (Benjamini & Hochberg, 1997). Weighted procedures are also needed in some applications even when the original hypotheses are equally weighted, as in gatekeeping procedures (Dmitrienko et al., 2007).

Suppose that there are $n \geqslant 2$ hypotheses, $H_1, \ldots, H_n$, which are to be tested based on their p-values, $p_1, \ldots, p_n$. Further suppose that positive weights, $w_1, \ldots, w_n$, which sum to 1, are preassigned to the hypotheses. Throughout we assume the standard requirement (Hochberg & Tamhane, 1987, p. 3) for any multiple test procedure that the familywise error rate,

$$\text{FWER} = \text{pr(Reject at least one true hypothesis)},$$

be controlled strongly, i.e., under all possible configurations of the true and false hypotheses, at a specified level α.

In the unweighted case, with all $w_i = 1/n$, the Hochberg and Holm procedures, which we shall denote by HC and HM, respectively, operate as follows. Let $p_{(1)} \leqslant \cdots \leqslant p_{(n)}$ be the ordered p-values and let $H_{(1)}, \ldots, H_{(n)}$ be the corresponding hypotheses. Procedure HC tests the hypotheses in a step-up manner starting with $H_{(n)}$, accepting one hypothesis at a time, and stopping at the ith step and rejecting $H_{(n-i+1)}, \ldots, H_{(1)}$ if $p_{(n-i+1)} \leqslant \alpha/i$. On the other hand, procedure HM tests the hypotheses in a step-down manner starting with $H_{(1)}$, rejecting one hypothesis at a time, and stopping at the ith step and accepting $H_{(i)}, \ldots, H_{(n)}$ if $p_{(i)} > \alpha/(n - i + 1)$. Since both procedures use the same critical constants, it is clear that HC rejects all the hypotheses rejected by HM, and possibly more.

For guaranteed familywise error rate control by procedure HC, the p-values must be independent since that assumption underlies the Simes (1986) test on which it is based. This requirement can be relaxed to allow positively dependent p-values using the extension of the Simes test to this case by Sarkar & Chang (1997) and Sarkar (1998). Procedure HM and the Bonferroni test on which it is based do not make the independence assumption. In most of this paper we assume that the p-values are independent; the dependent case is discussed in § 5.

Two types of weighted procedures have been proposed in the literature. Type-1 procedures are based on ordered raw p-values, p_i, while Type-2 procedures are based on ordered weighted p-values, $p_i^* = p_i/w_i$. We denote the corresponding weighted Holm procedures by WHM1 and WHM2, and weighted Simes procedures by WSM1 and WSM2, respectively. Note that WSM1 and WSM2 are global, not multiple tests, which we use in this paper to construct two weighted closed procedures, WCL1 and WCL2. We also study two weighted Hochberg procedures, denoted by WHC1 and WHC2, that are step-up counterparts of WHM1 and WHM2, respectively. We compare these procedures based on the criteria of the familywise error rate control and power. Another criterion that we use to compare different weighted Hochberg procedures is their monotonicity in terms of p-values: if one or more p-values are made smaller while the other p-values are kept unchanged then a procedure should reject at least the same or more hypotheses. Although intuitively plausible, some weighted procedures do not satisfy this property, as shown by Benjamini & Hochberg (1997) for WHM1.

2. WEIGHTED CLOSED PROCEDURES

2·1. *Type-1 weighted closed procedure,* WCL1

The original Hochberg procedure was derived as a conservative approximation to a closed procedure in which the Simes (1986) global test was used to test all subset intersection hypotheses. We follow the same approach in the weighted case. In this and the following subsection, we construct weighted closed procedures, WCL1 and WCL2, using the weighted Simes tests, WSM1 and WSM2, respectively. The WSM1 test used in the weighted closed procedure below is based on the following generalization of the Simes identity due to Benjamini & Hochberg (1997). If $p_{(1)} \leqslant \cdots \leqslant p_{(n)}$ are order statistics of n independent and identically distributed Un(0, 1) random variables then

$$\text{pr}\left\{p_{(i)} > \left(\sum_{k=1}^{i} w_{(k)}\right)\alpha \ \text{ for all } \ i = 1, \ldots, n\right\} = 1 - \alpha. \tag{1}$$

Procedure WCL1 operates as follows. Let $I = \{i_1, \ldots, i_m\}$ be a nonempty subset of the index set $N = \{1, \ldots, n\}$. Furthermore, let $p_{(i_1)} \leqslant \cdots \leqslant p_{(i_m)}$ be the ordered p-values and let $w_{(i_1)}, \ldots, w_{(i_m)}$ be the weights associated with them. Throughout, it will be assumed that ties between the p-values are broken at random. Then test and reject $H_I = \bigcap_{j=1}^{m} H_{(i_j)}$ at level α if and only if all H_J for $J \supset I$ are rejected and

$$p_{(i_j)} \leqslant \frac{\sum_{k=1}^{j} w_{(i_k)}}{\sum_{k=1}^{m} w_{(i_k)}}\alpha, \ \text{for at least one } j = 1, \ldots, m. \tag{2}$$

Since this is an α-level test of every intersection hypothesis, it follows from the closure principle of Marcus et al. (1976) that procedure WCL1 controls the familywise error rate at the α-level. The following example illustrates this procedure.

Example 1. The table below gives the p-values and weights for three hypotheses.

H_1	H_2	H_3
$p_1 = 0{\cdot}03$	$p_2 = 0{\cdot}035$	$p_3 = 0{\cdot}1$
$w_1 = 0{\cdot}2$	$w_2 = 0{\cdot}6$	$w_3 = 0{\cdot}2$

Procedure WCL1 begins by testing $H_1 \cap H_2 \cap H_3$ and rejects it since $p_2 \leqslant (w_1 + w_2)\alpha = 0{\cdot}04$.

Next it tests $H_1 \cap H_2$, $H_1 \cap H_3$ and $H_2 \cap H_3$. Both intersections that include H_2 are rejected since, in the case of $H_1 \cap H_2$, $p_2 \leqslant \alpha = 0{\cdot}05$ and, in the case of $H_2 \cap H_3$, $p_2 \leqslant \{w_2/(w_2 + w_3)\}\alpha = 0{\cdot}0375$. However, $H_1 \cap H_3$ cannot be rejected since $p_1 > \{w_1/(w_1 + w_3)\}\alpha = 0{\cdot}025$ and $p_3 > \alpha = 0{\cdot}05$. Therefore, in the last stage, we only test H_2 and reject it since $p_2 \leqslant \alpha = 0{\cdot}05$. Thus, WCL1 accepts H_1 and H_3, but rejects H_2.

Procedure WCL1 lacks a simple stepwise structure, so it is not easy to apply it by hand for larger n, although it is easy to program. Also, it is not easy to explain it to practitioners. To overcome these drawbacks, we next derive a conservative step-up approximation to it. We first introduce the idea of critical matrix for a closed procedure, due to Liu (1996). Assume that the hypotheses are equally weighted and there exist critical constants $c_{m1} \geqslant \cdots \geqslant c_{mm}$ for testing any intersection hypothesis $H_I = \bigcap_{j=1}^{m} H_{(i_j)}$ for a nonempty subset $I \subseteq N$ such that, under H_I,

$$\min_{H_I} \text{pr}\{p_{(i_1)} > c_{mm}, \ldots, p_{(i_m)} > c_{m1}\} \geqslant 1 - \alpha, \tag{3}$$

where the minimum is taken over all H_I of given cardinality m, where $m = 1, \ldots, n$. Then the test that rejects H_I if

$$p_{(i_j)} \leqslant c_{m,m-j+1} \text{ for at least one } j = 1, \ldots, m$$

has level α. Hence the closed procedure that uses this test for all intersection hypotheses controls the familywise error rate at level α. The Simes identity is a special case of (3) for $c_{mj} = (m - j + 1)\alpha/m$.

The critical matrix is a lower-triangular matrix of which the mth row gives the ordered critical constants for testing any intersection hypothesis of cardinality m and is given by

$$C = \begin{bmatrix} c_{11} & & & \\ c_{21} & c_{22} & & \\ \vdots & \vdots & \ddots & \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{bmatrix}.$$

Theorem 2 of Liu (1996) shows that if the elements of each row of C are equal then the closed procedure has a step-down short-cut. For example, if $c_{mj} = \alpha/m$ ($j = 1, \ldots, m$), namely the Bonferroni critical constants, then we obtain procedure HM; in that case, the first column of C from the bottom up, namely $(\alpha/n, \alpha/(n-1), \ldots, \alpha)$, gives the critical constants for testing the ordered p-values, $p_{(1)} \leqslant \cdots \leqslant p_{(n)}$. Theorem 3 of Liu (1996) shows that if the elements of each column of C are equal then the closed procedure has a step-up short-cut. For example, if $c_{mj} = \alpha/j$ ($j = 1, \ldots, m$) then we obtain procedure HC; in that case, the last row, namely $(\alpha, \alpha/2, \ldots, \alpha/n)$, gives the critical constants for testing the ordered p-values, $p_{(n)} \geqslant \cdots \geqslant p_{(1)}$. If one uses the Simes critical constants, $c_{mj} = (m - j + 1)\alpha/m$, which are sharper than the constants in procedure HC, $c_{mj} = \alpha/j$, then we get Hommel's (1988) procedure; in this case C does not have equal column entries and hence lacks a simple step-up short-cut. Rom (1990) provided a method for computing the critical constants c_{mj} under the constraint $c_{mj} = c_{m+1,j}$, which makes the column entries equal. Rom used the usual ordering of the critical constants, namely $c_{m1} \leqslant \cdots \leqslant c_{mm}$, and the constraint $c_{mj} = c_{m+1,j+1}$, thus making the subdiagonal entries equal, which is equivalent to making the column entries of C equal in Liu's notation.

If the hypotheses are not equally weighted, the critical constants of procedure WCL1 are generally different for each subset of size m, except for $m = 1$, there being ${}_nC_m$ rows of critical constants for each m, or $(2^n - 1)$ rows altogether; thus the critical matrix is not square. For $n = 3$, this matrix is given by

$$\begin{array}{cc} C = \begin{bmatrix} \alpha & & \\ \alpha & & \\ \alpha & & \\ \alpha & \{w_{(1)}/(w_{(1)} + w_{(2)})\}\alpha & \\ \alpha & \{w_{(1)}/(w_{(1)} + w_{(3)})\}\alpha & \\ \alpha & \{w_{(2)}/(w_{(2)} + w_{(3)})\}\alpha & \\ \alpha & (w_{(1)} + w_{(2)})\alpha & w_{(1)}\alpha \end{bmatrix} & \begin{array}{c} H_{(1)} \\ H_{(2)} \\ H_{(3)} \\ H_{(1)} \cap H_{(2)} \\ H_{(1)} \cap H_{(3)} \\ H_{(2)} \cap H_{(3)} \\ H_{(1)} \cap H_{(2)} \cap H_{(3)}. \end{array} \end{array}$$

The intersection hypothesis tested by each row of critical constants is shown alongside that row.

To obtain a step-up short-cut to procedure WCL1, we use Theorem 3 of Liu (1996) and make the column entries of C equal by replacing the entries in each column by their minima. The resulting closed procedure is obviously conservative; we denote it by CWCL1. The first column has all entries equal to α, so the minimum is α. In the second column, the entries $\{w_{(1)}/(w_{(1)} + w_{(3)})\}\alpha$ and $\{w_{(2)}/(w_{(2)} + w_{(3)})\}\alpha$ are both less than or equal to the bottom entry, $(w_{(1)} + w_{(2)})\alpha$. Therefore, the second column can be replaced by the minimum of the top three entries. Finally, the third column has only one entry. Let

$$\gamma_1 = 1, \gamma_2 = \min\left\{\frac{w_{(1)}}{w_{(1)} + w_{(2)}}, \frac{w_{(1)}}{w_{(1)} + w_{(3)}}, \frac{w_{(2)}}{w_{(2)} + w_{(3)}}\right\}, \gamma_3 = w_{(1)}.$$

We reject an intersection hypothesis $H_I = \bigcap_{j=1}^m H_{(i_j)}$ if $p_{(i_j)} \leqslant \gamma_{m-j+1}\alpha$ for some $j = 1, \ldots, m$. Thus any intersection hypothesis of cardinality m uses the same set of critical constants that depend only on m. Therefore the ${}_nC_m$ rows of critical constants for all intersection hypotheses of cardinality m become equal and can be collapsed into a single row, resulting in the square critical matrix

$$C = \begin{bmatrix} \gamma_1\alpha & & \\ \gamma_1\alpha & \gamma_2\alpha & \\ \gamma_1\alpha & \gamma_2\alpha & \gamma_3\alpha \end{bmatrix}.$$

The last row gives the sequence of the critical constants, $(\gamma_1\alpha, \gamma_2\alpha, \gamma_3\alpha)$, for procedure CWCL1 for testing $p_{(3)} \geqslant p_{(2)} \geqslant p_{(1)}$.

In general, it can be shown that

$$\gamma_1 = 1,$$

$$\gamma_2 = \min_{(i_1),(i_2)} \left\{ \frac{w_{(i_1)}}{w_{(i_1)} + w_{(i_2)}} \right\},$$

$$\gamma_3 = \min_{(i_1),(i_2),(i_3)} \left\{ \frac{w_{(i_1)}}{w_{(i_1)} + w_{(i_2)} + w_{(i_3)}} \right\},$$

$$\vdots$$

$$\gamma_n = w_{(1)},$$

where $w_{(i_1)}, \ldots, w_{(i_m)}$ are the weights associated with the ordered p-values, $p_{(i_1)} \leqslant \cdots \leqslant p_{(i_m)}$, for any nonempty subset $I \subseteq N$. It follows from the definition that $\gamma_1 \geqslant \cdots \geqslant \gamma_n$ and hence the critical values, $c_i = \gamma_{n-i+1}\alpha$, are monotone: $c_1 \leqslant \cdots \leqslant c_n$. The procedure continues by testing the ordered p-values as long as $p_{(j)} > \gamma_{n-j+1}\alpha$ for $j = n, \ldots, i+1$. If $p_{(i)} \leqslant \gamma_{n-i+1}\alpha$ then it rejects $H_{(i)}, \ldots, H_{(1)}$ and stops testing. In the case of equal weights, we have $\gamma_m = 1/m$ ($m = 1, \ldots, n$) and we obtain procedure HC. It is easy to show that WCL1 and CWCL1 are identical when $n = 2$.

Example 2. Refer to the data from Example 1. We have

$$\gamma_1 = 1, \gamma_2 = \min\left\{ \frac{0{\cdot}2}{0{\cdot}2 + 0{\cdot}6}, \frac{0{\cdot}2}{0{\cdot}2 + 0{\cdot}2}, \frac{0{\cdot}6}{0{\cdot}6 + 0{\cdot}2} \right\} = 0{\cdot}25, \gamma_3 = 0{\cdot}2.$$

Hence CWCL1 compares $p_{(3)} = 0{\cdot}1$ with $\gamma_1\alpha = 0{\cdot}05$, so does not reject, $p_{(2)} = 0{\cdot}035$ with $\gamma_2\alpha = 0{\cdot}0125$, so does not reject, and $p_{(1)} = 0{\cdot}03$ with $\gamma_3\alpha = 0{\cdot}01$, so does not reject. Recall that WCL1 rejects H_2.

2·2. *Type-2 weighted closed procedure,* WCL2

Hochberg & Liberman (1994) proposed a Type-2 weighted Simes test, WSM2, based on weighted p-values, $p_i^* = p_i/w_i$. Let $p_{(1)}^* \leqslant \cdots \leqslant p_{(n)}^*$ be their ordered values. Let $H_{(i)}^*$ be the hypotheses and $w_{(i)}^*$ the weights associated with the ordered $p_{(i)}^*$. Procedure WSM2 requires that $\sum w_i = 1$ and $\max w_i \leqslant 1/n\alpha$. It rejects the overall intersection hypothesis $\bigcap_{i=1}^n H_i$ if

$$p_{(j)}^* \leqslant j\alpha \text{ for at least one } j = 1, \ldots, n. \quad (4)$$

In this section, we explore the possibility of basing a closed test procedure, WCL2, on WSM2.

Benjamini & Hochberg (1997) showed that WSM1 is more powerful than WSM2 if the weights are ordered opposite to that of the p-values; this would happen if higher weights were assigned to the hypotheses, which are expected to give the most significant results, and the a priori ordering of the weights turns out to be correct. In general, neither procedure dominates the other. However, using the WSM2 test in WCL2 leads to some practical difficulties and a rather unwieldy procedure as explained below.

Procedure WCL2 uses WSM2 to test any subset intersection hypothesis $H_I = \bigcap_{j=1}^m H_{i_j}$ as follows. Recalculate the weights for the hypotheses to normalize them:

$$w_{i_j}^{(I)} = \frac{w_{i_j}}{\sum_{k=1}^m w_{i_k}} \quad (j = 1, \ldots, m).$$

This normalization is not required for WSM1 since its critical constants are functions of the ratios of weights, so the normalizing constant cancels out. Furthermore, if any $w_{i_j}^{(I)} > 1/m\alpha$, then set it equal to $1/m\alpha$; this makes $\sum w_{i_j}^{(I)} < 1$ and the procedure becomes conservative. Recalculate the weighted p-values,

$$p_{i_j}^{*(I)} = \frac{p_{i_j}}{w_{i_j}^{(I)}} \quad (j = 1, \ldots, m),$$

and order them as $p_{(i_1)}^{*(I)} \leqslant \cdots \leqslant p_{(i_m)}^{*(I)}$. Then reject H_I at level α if and only if all H_J for $J \supset I$ are rejected and

$$p_{(i_j)}^{*(I)} \leqslant j\alpha \text{ for at least one } j = 1, \ldots, m.$$

As one can see, WCL2 is rather complicated because of the need to recalculate the weights and weighted p-values for every intersection hypothesis. The restriction that $\max w_{i_j}^{(I)} \leqslant 1/m\alpha$ can be severe if m is large. Also, it is not possible to give a simple conservative step-up short-cut to WCL2, as we were able to do for WCL1. Given these difficulties and the fact that WCL2 does not have a uniform power advantage over WCL1, we chose not to consider WCL2 any further.

3. STEP-UP COUNTERPARTS OF WEIGHTED HOLM PROCEDURES

3·1. *Type*-1 *weighted Hochberg procedure,* WHC1

Let $p_{(1)} \leqslant \cdots \leqslant p_{(n)}$ be the ordered p-values, and let $H_{(i)}$ be the hypotheses and $w_{(i)}$ the weights associated with the ordered $p_{(i)}$. Benjamini & Hochberg (1997) proposed a weighted Holm procedure, WHM1, which operates as follows. At the first step, test $H_{(1)}$ by comparing $p_{(1)}$ with $w_{(1)}\alpha$. If $p_{(1)} > w_{(1)}\alpha$ then accept all hypotheses and stop testing; otherwise, reject $H_{(1)}$ and proceed to test $H_{(2)}$ by comparing $p_{(2)}$ with $w_{(2)}\alpha / \sum_{k=2}^n w_{(k)}$. In general, accept $H_{(i)}, \ldots, H_{(n)}$ if

$$p_{(i)} > \frac{w_{(i)}}{\sum_{k=i}^n w_{(k)}} \alpha$$

and stop testing; otherwise reject $H_{(i)}$ and continue to test $H_{(i+1)}$.

By analogy with WHM1, we propose the following procedure WHC1. At the first step, test $H_{(n)}$ by comparing $p_{(n)}$ with α. If $p_{(n)} \leqslant \alpha$ then reject all hypotheses and stop testing; otherwise, accept $H_{(n)}$ and proceed to test $H_{(n-1)}$ by comparing $p_{(n-1)}$ with $w_{(n-1)}\alpha / \sum_{k=n-1}^n w_{(k)}$. In general, reject $H_{(i)}, \ldots, H_{(1)}$ if

$$p_{(i)} \leqslant \frac{w_{(i)}}{\sum_{k=i}^{n} w_{(k)}} \alpha \qquad (5)$$

and stop testing; otherwise accept $H_{(i)}$ and continue to test $H_{(i-1)}$. Since WHC1 uses the same critical values as WHM1, as with the relationship between HM and HC, WHC1 is more powerful than WHM1.

Remark 1. If we can show that any hypothesis rejected by WHC1 is also rejected by WCL1 then it will follow that WHC1 controls the familywise error rate conservatively since WCL1 does. Unfortunately, this result is not true in general except when $n = 2$, in which case WCL1 and WHC1 are equivalent, or when the weights are equal, in which case Hochberg (1988) used this method of proof for procedure HC; however, there was a gap in his proof, which we fill in Remark 2. We now explain the only case where the proof using this method fails. The other cases where the method works are covered in a working paper available from the first author.

Suppose that, for $n \geqslant 2$, the largest i for which $p_{(i)}$ satisfies (5) is greater than 1 and less than n; so WHC1 rejects all $H_{(j)}$ for $j = 1, \ldots, i$. We will exhibit an intersection hypothesis H_I containing $H_{(j)}$ for $j < i$ such that WCL1 accepts H_I and hence also $H_{(j)}$. Let $I = \{(j), (n)\}$. Then WCL1 finds H_I significant at level α if

$$p_{(j)} \leqslant \frac{w_{(j)}}{w_{(j)} + w_{(n)}} \alpha \text{ or } p_{(n)} \leqslant \alpha.$$

We know that $p_{(n)} > \alpha$ and the first inequality is not always satisfied unless $w_{(j)} \geqslant w_{(i)}$, in which case

$$p_{(j)} \leqslant p_{(i)} \leqslant \frac{w_{(i)}}{w_{(i)} + w_{(n)}} \alpha \leqslant \frac{w_{(j)}}{w_{(j)} + w_{(n)}} \alpha. \qquad (6)$$

However, this condition cannot be guaranteed because of the randomness of the $w_{(j)}$ associated with the ordered $p_{(j)}$ values unless, of course, the weights are equal. Example 3 illustrates this point.

Remark 2. For equal weights, $w_i = 1/n$, we show that if $p_{(j)} \leqslant p_{(i)} \leqslant \alpha/(n - i + 1)$ then the unweighted closed procedure CL rejects all H_I containing $H_{(j)}$ for $j < i$, which Hochberg did not show. Suppose that H_I contains s hypotheses, including $H_{(j)}$, with p-values $\leqslant p_{(i)}$ and t hypotheses with p-values $> p_{(i)}$ and let $r \leqslant s$ be the rank of $p_{(j)}$ among the $s + t$ ordered p-values of these hypotheses. Then, in the test of H_I, CL compares $p_{(j)}$ with $\{r/(s + t)\}\alpha$. However,

$$p_{(j)} \leqslant p_{(i)} \leqslant \frac{1}{n - i + 1} \alpha \leqslant \frac{r}{s + t} \alpha,$$

where the last inequality follows since $r \geqslant 1$ and $s + t \leqslant n - i + 1$. Hence CL rejects all H_I containing $H_{(j)}$.

Example 3. Consider the data from Example 1. The critical constants for WHC1 are shown in the following table.

H_1	H_2	H_3
$p_1 = 0{\cdot}03$	$p_2 = 0{\cdot}035$	$p_3 = 0{\cdot}1$
$w_1 = 0{\cdot}2$	$w_2 = 0{\cdot}6$	$w_3 = 0{\cdot}2$
$c_1 = w_1\alpha = 0{\cdot}01$	$c_2 = \{w_2/(w_2 + w_3)\}\alpha = 0{\cdot}0375$	$c_3 = \alpha = 0{\cdot}05$

Procedure WHC1 accepts H_3 since $p_3 > 0{\cdot}05$, but rejects H_1 and H_2 since $p_2 < 0{\cdot}0375$. Recall that WCL1 does not reject H_1, so that WHC1 is not more conservative than WCL1. This is because

$$\frac{w_1}{w_1 + w_3}\alpha = \frac{1}{2}\alpha < \frac{w_2}{w_2 + w_3}\alpha = \frac{3}{4}\alpha,$$

which violates the required inequality (6) because $w_1 < w_2$. In this example, CWCL1 does not reject any hypothesis, nor does WHM1 since $p_1 = 0{\cdot}03 > c_1 = 0{\cdot}01$.

Next we show that a direct proof of the familywise error rate control of WHC1 along the lines of Liu (1996) also fails. This direct proof is based on showing that the probability of accepting all true hypotheses is minimized when the p-values from all false hypotheses approach zero, and the resulting lower bound equals $1 - \alpha$ from the Simes identity (1). To be specific, suppose that $H_1, \ldots, H_m$ are true and $H_{m+1}, \ldots, H_n$ are false. If we can show that the familywise error rate is maximized when $p_i \to 0$ $(i = m + 1, \ldots, n)$ then we would obtain

$$\begin{aligned} 1 - \text{FWER} &\geqslant \text{pr}\left\{ p_{(j)} > \frac{w_{(j)}}{\sum_{k=j}^{m} w_{(k)}}\alpha \text{ for all } j = 1, \ldots, m \right\} \\ &\geqslant \text{pr}\left\{ p_{(j)} > \frac{\sum_{k=1}^{j} w_{(k)}}{\sum_{k=1}^{m} w_{(k)}}\alpha \text{ for all } j = 1, \ldots, m \right\} \\ &= 1 - \alpha, \end{aligned}$$

where the second step follows because

$$\frac{w_{(j)}}{\sum_{k=j}^{m} w_{(k)}} \leqslant \frac{\sum_{k=1}^{j} w_{(k)}}{\sum_{k=1}^{m} w_{(k)}}$$

and the final step follows from (1). Unfortunately, it is not always true, except when the weights are equal, that familywise error rate is maximized when the p-values for all false hypotheses approach zero as shown by the following counterexample.

Example 4. Consider again the data from Example 1, and assume that H_1 and H_3 are true and H_2 is false. Then WHC1 rejects H_1 and H_2, thus committing a Type-1 error. Compare this scenario with the following in which we let $p_2 \to 0$. Then the p-values are ordered as $p_2 < p_1 < p_3$. The critical values to which they are compared equal

$$c_2 = w_2\alpha = 0{\cdot}03, \quad c_1 = \frac{w_1}{w_1 + w_3}\alpha = 0{\cdot}025, \quad c_3 = \alpha = 0{\cdot}05.$$

Procedure WHC1 rejects only H_2 in this case and hence does not commit a Type-1 error. Therefore letting $p_2 \to 0$ decreases the chance of making a Type-1 error. This is a consequence of the critical values not being monotone.

Despite these negative results, the simulations reported in § 4 fail to show a single case where WHC1 is anticonservative; in fact, in all the cases that we studied, WHC1 is slightly more conservative than WCL1. Thus, for practical purposes, WHC1 may be regarded as acceptable with regard to the familywise error rate control.

3·2. *Type-2 weighted Hochberg procedure,* WHC2

Holm (1979) proposed procedure WHM2, which operates as follows. Let $H^*_{(1)}, \ldots, H^*_{(n)}$ be the hypotheses associated with the ordered weighted p-values, $p^*_{(1)} \leqslant \cdots \leqslant p^*_{(n)}$. At the first step, test

$H^*_{(1)}$ by comparing $p^*_{(1)}$ with α. If $p^*_{(1)} > \alpha$ then accept all hypotheses and stop testing; otherwise, reject $H^*_{(1)}$ and proceed to test $H^*_{(2)}$ by comparing $p^*_{(2)}$ with $\alpha / \sum_{k=2}^{n} w^*_{(k)}$. In general, accept $H^*_{(i)}, \ldots, H^*_{(n)}$ if

$$p^*_{(i)} > \frac{\alpha}{\sum_{k=i}^{n} w^*_{(k)}} \tag{7}$$

and stop testing; otherwise reject $H^*_{(i)}$ and continue to test $H^*_{(i+1)}$.

By analogy with WHM2, we consider procedure WHC2, which operates as follows. At the first step, test $H^*_{(n)}$ by comparing $p^*_{(n)}$ with $\alpha / w^*_{(n)}$, i.e., by comparing the p_i corresponding to $p^*_{(n)}$ with α. If that $p_i \leqslant \alpha$ then reject all hypotheses and stop testing; otherwise, accept $H_i = H^*_{(n)}$ and proceed to test $H^*_{(n-1)}$ by comparing $p^*_{(n-1)}$ with $\alpha / \sum_{k=n-1}^{n} w^*_{(k)}$. In general, reject $H^*_{(i)}, \ldots, H^*_{(1)}$ if

$$p^*_{(i)} \leqslant \frac{\alpha}{\sum_{k=i}^{n} w^*_{(k)}} \tag{8}$$

and stop testing; otherwise accept $H^*_{(i)}$ and continue to test $H^*_{(i-1)}$. Unfortunately, WHC2 does not control the familywise error rate, even though WHM2 does, as shown in the following proposition for $n = 2$. Simulations show that lack of familywise error rate control persists for $n > 2$, and therefore we drop procedure WHC2 from consideration.

PROPOSITION 1. *For $n = 2$, under the overall null hypothesis $H = H_1 \cap H_2$, the familywise error rate for* WHC2 *is symmetric about $w_1 = 1/2$ and, for $w_1 \leqslant 1/2$, is given by*

$$\text{FWER} = \begin{cases} \alpha(1-\alpha) + \alpha^2\,(w_1/w_2 + w_2/w_1)/2, & \text{if } \alpha/(1+\alpha) \leqslant w_1 \leqslant 1/2, \\ 1 - (1-\alpha)^2 - (w_1/2w_2)(1-\alpha^2), & \text{if } w_1 \leqslant \alpha/(1+\alpha). \end{cases} \tag{9}$$

The minimum of this function is attained when $w_1 = w_2 = 1/2$ and equals α. The maximum is attained when $w_1 = 0$ or $w_1 = 1$ and equals $1 - (1-\alpha)^2$. Therefore, FWER $\geqslant \alpha$ *for all (w_1, w_2) with equality holding if and only if $w_1 = w_2 = 1/2$.*

The proof is given in the Appendix.

Figure 1 shows the plot of the familywise error rate of WHC2 given by (9) for $\alpha = 0{\cdot}05$. For $w_1 = 0{\cdot}5$, FWER $= 0{\cdot}05$ and, for $w_1 \neq 0{\cdot}5$, FWER $> 0{\cdot}05$ approaching $1 - (0{\cdot}95)^2 = 0{\cdot}0975$ as $w_1 \to 0$ or 1.

4. SIMULATION RESULTS FOR INDEPENDENT p-VALUES

4·1. *Results about the familywise error rate*

We carried out simulations to study (i) whether or not WHC1 in fact controls the familywise error rate, (ii) how closely WCL1 controls the familywise error rate, and (iii) how conservative CWCL1 is. We first focused on the case of $n = 3$ and studied a range of configurations of true and false hypotheses, and the associated weights. The 'true' p-values were generated from the Un(0, 1) distribution while the 'false' p-values were generated by first generating z-statistics from a $N(\delta, 1)$ distribution with $\delta = 1{\cdot}0$ and calculating their one-sided p-values from the formula $p = \text{pr}\{N(0, 1) \geqslant z\} = 1 - \Phi(z)$, where $\Phi(z)$ is the standard normal cumulative distribution function.

For $n = 3$, we studied the three cases of $m = 1, 2$ or 3 true hypotheses. Since the familywise error rates of all three procedures, WHC1, WCL1 and CWCL1, were highest when all three hypotheses were true, the results only for that case are reported in Table 1. Each estimate of

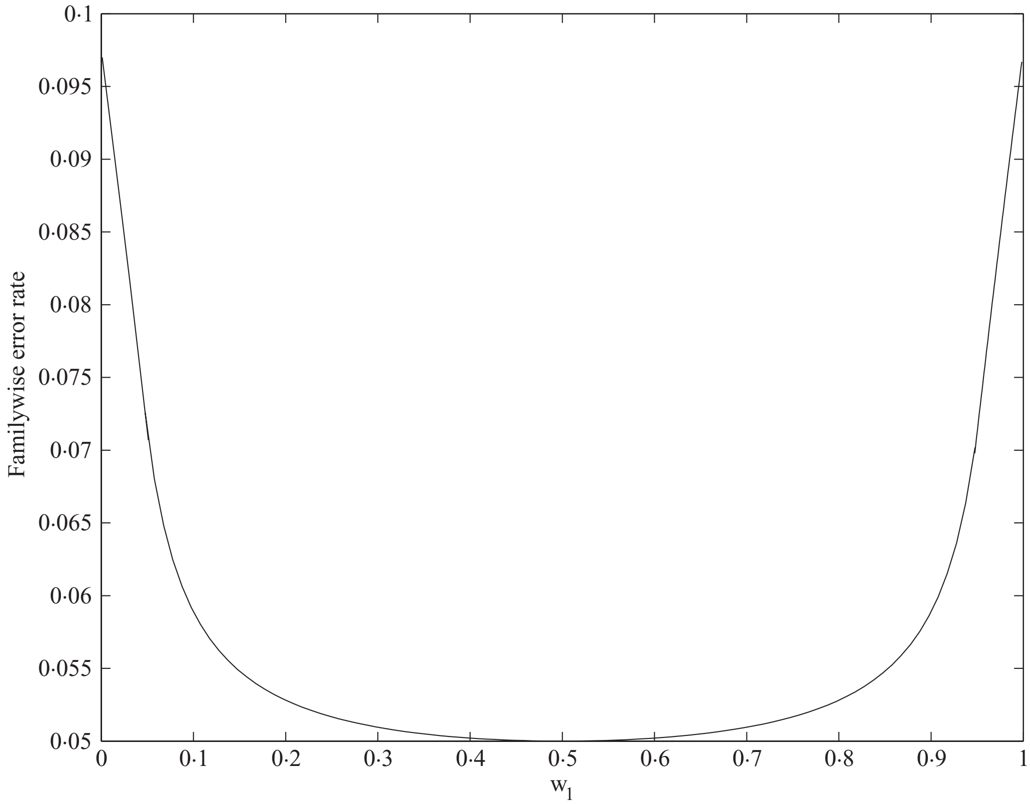

Fig. 1. Plot of the familywise error rate for procedure WHC2 for $n = 2$ under the overall null hypothesis ($\alpha = 0{\cdot}05$).

Table 1. *Estimates of familywise error rate for independent p-values. Here $n = 3$, and all three hypotheses are true*

Weights (T, T, T)	WHC1	CWCL1	WCL1
(0·1, 0·45, 0·45)	0·0491	0·0485	0·0492
(0·2, 0·4, 0·4)	0·0498	0·0496	0·0502
(0·3, 0·35, 0·35)	0·0488	0·0488	0·0491
(0·4, 0·3, 0·3)	0·0493	0·0492	0·0496
(0·5, 0·25, 0·25)	0·0490	0·0485	0·0494
(0·6, 0·2, 0·2)	0·0495	0·0487	0·0496
(0·7, 0·15, 0·15)	0·0489	0·0478	0·0490
(0·8, 0·1, 0·1)	0·0498	0·0485	0·0499
(0·9, 0·05, 0·05)	0·0496	0·0478	0·0496

T, true hypothesis.

the familywise error rate is based on a total of 100 000 replications. The deviations of the estimates from the nominal value of $\alpha = 0{\cdot}05$ should be standardized by the standard error, $\surd\{0{\cdot}05 \times 0{\cdot}95/100\,000\} = 0{\cdot}0007$, in order to determine whether or not the estimates differ significantly from α.

We see that WCL1 controls the familywise error rate very accurately and WHC1 is always slightly conservative compared to WCL1. In no case does the estimated familywise error rate of WHC1 exceed more than two standard deviations above the nominal value, i.e., higher than

Table 2. *Estimates of power for independent p-values with $n = 3$, $\delta = 2$, one true and two false hypotheses*

Weights (T, F, F)	WHM1	WHC1	CWCL1	WCL1
(0·1, 0·45, 0·45)	0·7469	0·7595	0·7484	0·7627
(0·2, 0·4, 0·4)	0·7273	0·7368	0·7298	0·7424
(0·3, 0·35, 0·35)	0·7049	0·7116	0·7099	0·7182
(0·4, 0·3, 0·3)	0·6781	0·6837	0·6833	0·6906
(0·5, 0·25, 0·25)	0·6445	0·6501	0·6486	0·6555
(0·6, 0·2, 0·2)	0·6072	0·6133	0·6110	0·6181
(0·7, 0·15, 0·15)	0·5576	0·5651	0·5614	0·5680
(0·8, 0·1, 0·1)	0·4855	0·4970	0·4907	0·4981
(0·9, 0·05, 0·05)	0·3788	0·3955	0·3862	0·3950
(0·45, 0·1, 0·45)	0·6160	0·6470	0·6215	0·6482
(0·4, 0·2, 0·4)	0·6653	0·6780	0·6698	0·6823
(0·35, 0·3, 0·35)	0·6912	0·6976	0·6964	0·7043
(0·3, 0·4, 0·3)	0·7043	0·7122	0·7078	0·7187
(0·25, 0·5, 0·25)	0·7050	0·7192	0·7073	0·7241
(0·2, 0·6, 0·2)	0·6994	0·7243	0·7014	0·7273
(0·15, 0·7, 0·15)	0·6836	0·7230	0·6861	0·7244
(0·1, 0·8, 0·1)	0·6605	0·7187	0·6635	0·7194
(0·05, 0·9, 0·05)	0·6150	0·7052	0·6201	0·7053

T, true hypothesis, F, false hypothesis.

$0{\cdot}05 + (2)(0{\cdot}0007) = 0{\cdot}0514$. Thus, within the limits of sampling error, we conclude that WHC1 controls the familywise error rate. Procedure CWCL1 is, of course, conservative by construction. We performed additional simulations for $n = 5$, which are not reported here. They confirm that the familywise error rate of WHC1 is maximum when all hypotheses are true and this maximum does not significantly exceed the nominal $\alpha = 0{\cdot}05$.

We also performed simulations to study the familywise error rate of WHC2 for $n = 3$ when all three hypotheses are true. The results, not shown here, confirm that when the weight on one of the hypotheses approaches 0 or 1, the familywise error rate exceeds the nominal α by a significant amount. This result is in accord with the result for $n = 2$ in Proposition 1. The conclusion is that WHC2 is not recommended for use.

4·2. *Power results*

It is clear that, by construction, WHC1 is more powerful than WHM1. Similarly, WCL1 is more powerful than CWCL1. To assess the power advantages of these procedures over their conservative counterparts, as well as to compare the powers of WCL1 and WHC1, we performed another simulation study. We used the following definition of power.

$$\text{Power} = \text{pr(Reject at least one false hypothesis)}.$$

Only the case $n = 3$, with one true and two false hypotheses, is reported here. We used the same weight configurations as in § 4·1, and 100 000 replications were made for each run. The δ-values for the false hypotheses were set equal to 2·0. The results are shown in Table 2. We also performed simulations for an alternative definition of power, namely, pr(Reject all false hypotheses), but do not report the results here.

Clearly, WCL1 is more powerful than the other three procedures in all cases except for one extreme case, in which both false hypotheses have a weight of 0·05, with WHC1 a close second. The difference between the powers of WCL1 and WHC1 is not statistically significant in any case.

Table 3. *Estimates of familywise error rate for correlated p-values. Here $n = 3$, and all three hypotheses are true, with $\rho = 0{\cdot}5$*

Weights (T, T, T)	WHC1	CWCL1	WCL1
(0·1, 0·45, 0·45)	0·0464	0·0443	0·0468
(0·2, 0·4, 0·4)	0·0439	0·0430	0·0447
(0·3, 0·35, 0·35)	0·0436	0·0435	0·0447
(0·4, 0·3, 0·3)	0·0447	0·0442	0·0459
(0·5, 0·25, 0·25)	0·0446	0·0434	0·0454
(0·6, 0·2, 0·2)	0·0475	0·0451	0·0480
(0·7, 0·15, 0·15)	0·0462	0·0431	0·0465
(0·8, 0·1, 0·1)	0·0476	0·0432	0·0477
(0·9, 0·05, 0·05)	0·0495	0·0438	0·0496

T, true hypothesis.

For the alternative definition of power given in the previous paragraph, this was not necessarily the case and WHC1 had higher power than WCL1 in many cases, although the differences were again not statistically significant. The power of WHM1 is lowest in all cases. Both WHM1 and CWCL1 suffer their greatest statistically significant power loss compared to WCL1 and WHC1 when the weight on one of the false hypotheses is high, at least 0·6.

5. SIMULATION RESULTS FOR CORRELATED p-VALUES

Thus far we have assumed that the p-values are independent, but in practice the test statistics are often correlated, as for example in the case of multiple endpoints or comparisons between doses and a control. Based on the results of Sarkar (1998) and Sarkar & Chang (1997) regarding the conservativism of the Simes test under positive dependence in the unweighted case, one may conjecture that a similar result may hold true for the WSM1 test. This was proved by Y. Kling in his unpublished 2005 Ph.D. dissertation from Tel Aviv University. In that case, WCL1, and hence CWCL1, will control the familywise error rate under positive dependence; since WHC1 is found to be slightly more conservative than WCL1 for independent p-values, it is also likely to control the familywise error rate under positive dependence. By the same token, since the Simes test can be anticonservative under negative dependence (Samuel-Cahn, 1996), the same result may extend to WSM1, and hence WCL1 and WHC1 may be anticonservative under negative dependence. We investigated these conjectures via simulation.

We studied how the familywise error rates of WHC1 and WCL1 depend on the correlation, assumed to be common, between the p-values for the $n = 3$ case. One-sided p-values were imputed from equicorrelated normally distributed z-statistics with common correlation $\rho = 0{\cdot}1(0{\cdot}1)0{\cdot}9$ and also $\rho = -0{\cdot}4$; we must have $\rho > -0{\cdot}5$ in order for the joint distribution to be nondegenerate. Independent, normally distributed z-statistics were transformed by application of the Cholesky decomposition of the correlation matrix to obtain correlated z-statistics. The z-statistics from the true hypotheses were distributed as $N(0, 1)$ while those from the false hypotheses were $N(\delta, 1)$ with $\delta = 1{\cdot}0$. For positive correlations, we studied three cases with the number of true hypotheses $m = 1, 2, 3$. Keeping everything else fixed, we found that, for all three procedures, the familywise error rate decreases as ρ increases for $\rho < 1$. Similarly, the familywise error rate increases with m and the maximum is reached when $m = 3$, i.e., when all hypotheses are true. To save space, we report the results only for $m = 3$ in Table 3 for $\rho = 0{\cdot}5$ and in Table 4 for $\rho = -0{\cdot}4$.

Table 4. *Estimates of familywise error rate for correlated p-values. Here $n = 3$, and all three hypotheses are true, with $\rho = -0{\cdot}4$*

Weights (T, T, T)	WHC1	CWCL1	WCL1
(0·1, 0·45, 0·45)	0·0502	0·0501	0·0502
(0·2, 0·4, 0·4)	0·0489	0·0489	0·0490
(0·3, 0·35, 0·35)	0·0502	0·0502	0·0502
(0·4, 0·3, 0·3)	0·0507	0·0507	0·0508
(0·5, 0·25, 0·25)	0·0507	0·0506	0·0507
(0·6, 0·2, 0·2)	0·0504	0·0504	0·0504
(0·7, 0·15, 0·15)	0·0496	0·0494	0·0496
(0·8, 0·1, 0·1)	0·0488	0·0486	0·0488
(0·9, 0·05, 0·05)	0·0493	0·0491	0·0493

T, true hypothesis.

From these simulations, we see that for $\rho = 0{\cdot}5$ all three procedures control the familywise error rate conservatively. For $\rho = -0{\cdot}4$, the estimated familywise error rate values tend to be higher than those for $\rho = 0$, but none of them is significantly higher than the nominal $\alpha = 0{\cdot}05$. We checked whether or not the WSM1 test controls the Type-1 error rate for testing the overall intersection null hypothesis for $n = 3$ and $\rho = -0{\cdot}4$, and found that the test is anticonservative only for the weight configuration $(w_1, w_2, w_3) = (0{\cdot}3, 0{\cdot}35, 0{\cdot}35)$. This weight configuration is close to the equal-weights configuration, in which case the Simes test is known to be anticonservative, but the extent of anticonservatism is quite small. One may therefore conclude that any violation of the familywise error rate control in the case of negative correlations would occur when the weights are nearly equal, but this violation is likely to be small and therefore difficult to detect via simulation.

6. Monotonicity of procedures in terms of p-values

The notion of p-value monotonicity mentioned in § 1 is defined as follows. Let $P = (p_1, \ldots, p_n)$ and $P' = (p'_1, \ldots, p'_n)$ be two arbitrary vectors of p-values such that $p_i \leqslant p'_i$ for all i with a strict inequality for some i. Then a procedure is p-value monotone if it rejects the same hypotheses for the vector P as it does for the vector P', and possibly more.

It can be shown that both WSM1 and WSM2 are p-value monotone as tests of a single intersection hypothesis, because their critical values are monotonically increasing. Therefore, the weighted closed procedures, WCL1 and WCL2, based on them are also p-value monotone. On the other hand, WHC1 is not p-value monotone except for $n = 2$, in which case it is equivalent to WCL1; we show this in the following example. Procedure WHM1 is not p-value monotone even for $n = 2$, as was shown by Benjamini & Hochberg (1997).

Example 5. The table below gives the p-values and weights for three hypotheses.

H_1	H_2	H_3
$p_1 = 0{\cdot}022$	$p_2 = 0{\cdot}023$	$p_3 = 0{\cdot}055$
$w_1 = 0{\cdot}2$	$w_2 = 0{\cdot}4$	$w_3 = 0{\cdot}4$
$c_1 = w_1\alpha = 0{\cdot}01$	$c_2 = w_2/(w_2 + w_3)\alpha = 0{\cdot}025$	$c_3 = \alpha = 0{\cdot}05$

We see that WHC1 accepts H_3, but rejects H_1 and H_2. Now suppose p_1 and p_3 are kept fixed, but p_2 is reduced to 0·021. Then the ordering of the p-values changes, resulting in the table on the following page.

$$\begin{array}{ccc} H_2 & H_1 & H_3 \\ p_2 = 0{\cdot}021 & p_1 = 0{\cdot}022 & p_3 = 0{\cdot}055 \\ w_2 = 0{\cdot}4 & w_1 = 0{\cdot}2 & w_3 = 0{\cdot}4 \\ c_2 = w_2\alpha = 0{\cdot}02 & c_1 = w_1/(w_1 + w_3)\alpha = 0{\cdot}0167 & c_3 = \alpha = 0{\cdot}05 \end{array}$$

In this case WHC1 accepts all three hypotheses, although the p-values are no larger than before. It can be checked that WCL1 rejects H_2 in both cases. As a result of lack of p-value monotonicity, we do not recommend WHC1.

Roth (1999) introduced another type of monotonicity, referred to as α-consistency, defined as follows. If a procedure rejects a certain set of hypotheses for a given α then it should reject the same hypotheses and possibly more for $\alpha' > \alpha$. All the procedures discussed here are α-consistent since their critical values are increasing functions of α.

Remark 3. To avoid the problem discussed in Remark 1, one could consider a conservative modification of WHC1 that only rejects $H_{(i)}$ under (5) and not all $H_{(j)}$ for $j < i$. This is not a stepwise procedure as it applies the test (5) separately to each $H_{(i)}$. It is more conservative than WCL1 and hence controls the familywise error rate. It is easy to see that it rejects all the hypotheses rejected by WHM1 and possibly more. However, it also fails the test of p-value monotonicity as can be checked from the above example. Hence it is not recommended.

7. DISCUSSION

We have highlighted the difficulties inherent in constructing weighted Hochberg-type procedures. The natural step-up analogues of the two weighted Holm procedures fail the test of p-value monotonicity or the familywise error rate control. It is not simple to apply the Type-2 weighted closed procedure, WCL2, in practice and, in any case, it is not more powerful than WCL1. Thus we are left with WCL1 or its conservative step-up short-cut, CWCL1, as the only feasible alternatives. We recommend WCL1, although it has the drawback of difficulty of interpretation to practitioners. If the latter is an important consideration then CWCL1 may be used instead, but with a consequent loss of power. Simulations indicate that both WCL1 and CWCL1 control the familywise error rate under positive dependence, and may only be slightly anticonservative under negative dependence.

ACKNOWLEDGEMENT

This research was partially supported by the National Heart, Lung and Blood Institute and the National Security Agency. We are grateful to two referees and Professor D. M. Titterington for extensive comments and suggestions that led to a thorough revision of the paper.

APPENDIX

Proof of Proposition 1

The symmetry of the familywise error rate function of WHC2 about $w_1 = 1/2$ is obvious. Hence, without loss of generality, we may assume that $w_1 \leqslant w_2$, i.e., $w_1 \leqslant 1/2$. Let

$$p^*_{(1)} = \min\left\{\frac{p_1}{w_1}, \frac{p_2}{w_2}\right\}, \quad p^*_{(2)} = \max\left\{\frac{p_1}{w_1}, \frac{p_2}{w_2}\right\}.$$

Then

$$1 - \text{FWER} = \text{pr}\left(p^*_{(1)} > \frac{\alpha}{w_{(1)} + w_{(2)}},\, p^*_{(2)} > \frac{\alpha}{w_{(2)}}\right)$$
$$= \text{pr}\left(\frac{p_1}{w_1} > \alpha,\, p_2 > \alpha,\, \frac{p_1}{w_1} \leqslant \frac{p_2}{w_2}\right) + \text{pr}\left(\frac{p_2}{w_2} > \alpha,\, p_1 > \alpha,\, \frac{p_1}{w_1} > \frac{p_2}{w_2}\right)$$
$$= P_1 + P_2,$$

say.

We first evaluate

$$P_1 = \int_\alpha^1 \int_{w_1\alpha}^{w_1 p_2/w_2} dp_1\, dp_2 = \int_\alpha^1 \left(\frac{w_1}{w_2} p_2 - w_1\alpha\right) dp_2 = \frac{w_1}{2w_2} - w_1\alpha - \frac{w_1}{2w_2}\alpha^2 + w_1\alpha^2.$$

To evaluate P_2, we consider two cases. First, if

$$\frac{w_2}{w_1}\alpha \leqslant 1, \quad \text{i.e., } \frac{\alpha}{1+\alpha} \leqslant w_1 \leqslant \tfrac{1}{2},$$

then

$$P_2 = P_{21} = \int_{w_2\alpha}^{w_2\alpha/w_1} \int_\alpha^1 dp_1\ dp_2 + \int_{w_2\alpha/w_1}^1 \int_{w_1 p_2/w_2}^1 dp_1\ dp_2$$
$$= (1-\alpha)\left(\frac{w_2}{w_1}\alpha - w_2\alpha\right) + \left(1 - \frac{w_1}{2w_2} - \frac{w_2}{w_1}\alpha + \frac{w_2}{2w_1}\alpha^2\right)$$

and

$$1 - \text{FWER} = P_1 + P_{21} = 1 - \alpha + \alpha^2 - \frac{\alpha^2}{2}\left(\frac{w_1}{w_2} + \frac{w_2}{w_1}\right).$$

Secondly, if

$$\frac{w_2}{w_1}\alpha > 1, \quad \text{i.e., } w_1 < \frac{\alpha}{1+\alpha},$$

then

$$P_2 = P_{22} = \int_{w_2\alpha}^1 \int_\alpha^1 dp_1\ dp_2 = (1 - w_2\alpha)(1-\alpha)$$

and

$$1 - \text{FWER} = P_1 + P_{22} = (1-\alpha)^2 + \frac{w_1}{2w_2}(1-\alpha^2).$$

The expressions in (9) follow immediately. The minimum and maximum familywise error rates are α and $1 - (1-\alpha)^2 = \alpha(2-\alpha)$, respectively. This completes the proof of the proposition. □

References

Benjamini, Y. & Hochberg, Y. (1997). Multiple hypothesis testing with weights. *Scand. J. Statist.* **24**, 407–18.

Dmitrienko A., Wiens B. L., Tamhane A. C. & Wang X. (2007). Tree-structured gatekeeping tests in clinical trials with hierarchically ordered multiple objectives. *Statist. Med.* **26**, 2465–78.

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

Hochberg, Y. & Liberman, U. (1994). An extended Simes test. *Statist. Prob. Lett.* **21**, 101–5.

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

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.

LIU, W. (1996). 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. (1990). A sequentially rejective test procedure based on a modified Bonferroni inequality. *Biometrika* **77**, 663–5.

ROTH, A. (1999). Multiple comparison procedures for discrete test statistics. *J. Statist. Plan. Infer.* **82**, 101–17.

SAMUEL-CAHN, E. (1996). Is the Simes improved Bonferroni procedure conservative? *Biometrika* **83**, 928–33.

SARKAR, S. (1998). Some probability inequalities for the ordered MTP_2 random variables: A proof of the Simes conjecture. *Ann. Statist.* **26**, 494–504.

SARKAR, S. & CHANG, C. K. (1997). 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* **63**, 655–60.

[*Received March* 2007. *Revised December* 2007]