



A Step-Up Multiple Test Procedure

Charles W. Dunnett & Ajit C. Tamhane

To cite this article: Charles W. Dunnett & Ajit C. Tamhane (1992) A Step-Up Multiple Test Procedure, Journal of the American Statistical Association, 87:417, 162-170

To link to this article: <http://dx.doi.org/10.1080/01621459.1992.10475188>



Published online: 27 Feb 2012.



Submit your article to this journal [↗](#)



Article views: 40



View related articles [↗](#)



Citing articles: 83 View citing articles [↗](#)

A Step-Up Multiple Test Procedure

CHARLES W. DUNNETT and AJIT C. TAMHANE*

We consider the problem of simultaneously testing $k \geq 2$ hypotheses on parameters $\theta_1, \dots, \theta_k$. In a typical application the θ 's may be a set of contrasts, for instance, a set of orthogonal contrasts among population means or a set of differences between k treatment means and a standard treatment mean. It is assumed that least squares estimators $\hat{\theta}_1, \dots, \hat{\theta}_k$ are available that are jointly normally distributed with a common variance (known up to a scalar, namely the error variance σ^2) and a common known correlation. An independent χ^2 -distributed unbiased estimator of σ^2 is also assumed to be available. We propose a step-up multiple test procedure for this problem which tests the t statistics for the k hypotheses in order starting with the least significant one and continues as long as an acceptance occurs. (By contrast, the step-down approach, which is usually used, starts with the most significant and continues as long as a rejection occurs.) Critical constants required by this step-up procedure to control the type I familywise error rate at or below a specified level α are computed for both one-sided and two-sided testing problems. Power comparisons are made for one-sided testing problems with the well-known normal theory based single-step and step-down test procedures, and also with a step-up test procedure proposed for a wider class of problems by Hochberg. (Two improvements over Hochberg's procedure by Hommel and Rom provide at best marginal increases in power, with the former being also more difficult to apply, and hence they are not included here.) Two different definitions of power are considered—the probability of rejecting *all* false hypotheses and the probability of rejecting *at least one* false hypothesis; the results are found to be qualitatively similar. The proposed step-up procedure is more powerful than the single-step procedure except when only one hypothesis is false, in which case it is slightly less powerful. Similarly, it is slightly less powerful than the step-down procedure when a few hypotheses are false, but it is more powerful when most or all hypotheses are false, and this power advantage increases with the number of such hypotheses under test. The proposed step-up procedure is uniformly more powerful than the Hochberg procedure and its improvements. A disadvantage of the proposed step-up procedure is the greater difficulty of computing its critical points. These are given for one-sided and two-sided tests for 5% level of significance.

KEY WORDS: Comparisonwise approach; Experimentwise approach; Familywise error rate; Joint p -values; Multiple comparison procedures; Multivariate t distribution; Power; Single-step procedure; Step-down procedure.

1. INTRODUCTION

The purpose of many empirical studies is to compare several treatment groups by estimating differences or by performing tests of significance on relevant parameters. Such multiple comparisons in the same study are the rule rather than the exception. There has been much debate and controversy as to whether statistical adjustments are necessary for taking into account the multiplicity of inferences. One school of thought, as advanced by Miller (1966), Scheffé (1953), Tukey (1953), and others, is that such adjustments should be made by requiring the inference procedures to satisfy the condition that the probability of at least one wrong inference (e.g., at least one type I error in a multiple hypothesis testing problem) is controlled at or below a specified level α . This is called the "experimentwise" or the "familywise" error rate (FWE) control approach. The opposite school of thought, as represented, for example, by Carmer and Walker (1982), Nelder and some other discussants of O'Neill and Wetherill (1971), O'Brien (1983), Perry (1986), and Rothman (1990), maintains that such adjustments are not needed and that each inference should be dealt with separately, which is referred to as the "comparisonwise" error rate control approach. A third school of thought, led by Duncan [e.g., Duncan (1965), Waller and Duncan

(1969), Duncan and Dixon (1983)] adopts a Bayesian decision-theoretic approach to multiple comparisons problems assuming prior distributions on the unknown parameters, linear loss functions for type I and type II errors for individual significance tests, and an additive loss function for the overall loss. For the pairwise comparison problem, the resulting test procedure has the nature of a comparisonwise procedure in the sense that the critical constant it uses does not depend on the number of comparisons (as do the critical constants used by the familywise procedures); however, it does differ from the customary comparisonwise procedures [e.g., the unprotected least significant difference (LSD) test] by making this constant depend in an inverse way on the analysis of variance F statistic, which is a sample measure of the disparity between the treatment means being compared.

The use of the experimentwise approach is called for in the following situations: (1) when a conclusion is reached that requires the simultaneous correctness of several inferences and (2) when a conclusion is reached that hinges on the correctness of an inference that has been selected in light of the data. For example, consider a pharmaceutical company that compares several drugs with a standard in the same study for the purpose of determining which ones give the most improvement, with the goal of selecting one of them to recommend for further developmental work. If the drug that happens to be selected for this purpose is, in fact, inferior to the standard, then a serious error would be made. Therefore, it is necessary to control the probability of making such an error, which is achieved by using the experimentwise approach. On the other hand, if several treatments are included in the same experiment solely for the

* Charles W. Dunnett is Professor Emeritus, Department of Mathematics and Statistics and the Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada L8S 4K1. Ajit C. Tamhane is Professor, Department of Statistics and the Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL 60208. Charles W. Dunnett's research was supported by the Natural Sciences and Engineering Research Council of Canada. The authors thank Anthony J. Hayter for his help in investigating the existence of critical constants for the proposed procedure. The authors are also grateful to the two referees for their suggestions, which greatly improved the article.

purpose of economy and efficiency, and the comparisons of interest are inherently separate from each other and could logically have been examined in separate experiments were it not for the inefficiency of doing this, then the comparisonwise approach is appropriate (O'Brien 1983). Duncan's approach offers an alternative solution to multiple comparisons if the underlying assumptions, particularly the one concerning the additive form of the overall loss function (which is the reason for the comparisonwise nature of the procedure), are satisfied.

A criticism of the experimentwise approach sometimes made is that it lowers the power of detecting real effects due to the stricter requirement on the type I error rate. One way of increasing power, when inferences are in the form of hypotheses tests, is to employ stepwise tests. The purpose of this article is to develop a new stepwise procedure of the step-up type for testing multiple hypotheses under the framework of the experimentwise approach and compare its performance with a step-down procedure. We consider the problem of simultaneously testing $k \geq 2$ hypotheses, H_1, H_2, \dots, H_k , subject to the requirement that the type I FWE, which is the probability of rejecting any true hypothesis, be kept at or below a specified level α (referred to as the *overall significance level*). All procedures considered in this article satisfy this requirement. Some assume specific parametric distributional models, while others are more general. The prime example in this latter class is the well-known Bonferroni procedure, which rejects any H_i with $p_i \leq \alpha/k$, where p_i is the p value computed from the observed value t_i of some test statistic T_i for H_i ($1 \leq i \leq k$). For large k , the Bonferroni procedure becomes increasingly conservative, and hence modifications of it with improved powers have been proposed. One procedure due to Holm (1979) proceeds as follows: Order the p values, $p_{(1)} \geq \dots \geq p_{(k)}$, and denote the corresponding hypotheses, $H_{(1)}, \dots, H_{(k)}$. Start with the smallest p value, $p_{(k)}$. If $p_{(k)} > \alpha/k$, then stop testing and accept all the hypotheses; otherwise reject $H_{(k)}$ and go to the next step. In general, if testing has continued to the i th step ($1 \leq i \leq k$) and if $p_{(k-i+1)} > \alpha/(k-i+1)$, then stop testing and accept all the remaining hypotheses, $H_{(k-i+1)}, \dots, H_{(1)}$; otherwise reject $H_{(k-i+1)}$ and go to the next step. Another procedure due to Hochberg (1988) proceeds as follows: Start with the largest p value, $p_{(1)}$. If $p_{(1)} \leq \alpha$ then stop testing and reject all the hypotheses; otherwise accept $H_{(1)}$ and go to the next step. In general, if testing has continued to the i th step ($1 \leq i \leq k$) and if $p_{(i)} \leq \alpha/i$, then reject all the remaining hypotheses, $H_{(i)}, \dots, H_{(k)}$; otherwise accept $H_{(i)}$ and go to the next step. It is readily seen that the Holm procedure is uniformly less powerful than the Hochberg procedure, since the latter always rejects any hypothesis rejected by the former. This results from the fact that both use the same critical values for the $p_{(i)}$, but the Holm procedure uses a *step-down* algorithm, while the Hochberg procedure uses a *step-up* algorithm for making accept or reject decisions. Also note that the Holm procedure, in turn, is uniformly more powerful than the Bonferroni procedure, which is said to be of the *single-step* type (Hochberg and Tamhane 1987, chap. 2).

In the parametric case, normal theory based single-step and step-down procedures have been developed (described in Section 2.2), but hardly any attention has been paid to the step-up approach (the only exception to our knowledge being one proposed by Welsch (1977) for the pairwise comparisons problem). In this article, we introduce a new step-up procedure for the problem of testing a nonhierarchical family of hypotheses (Hochberg and Tamhane 1987, chap. 2) on $k \geq 2$ linear parametric functions (e.g., contrasts) and compare its performance with a competing step-down procedure. The development is restricted to balanced data situations, but extensions to unbalanced cases are covered briefly in Section 8.

2. PRELIMINARIES

2.1 Distributional Setup

We assume the standard normal theory linear model setting. Thus consider $k \geq 2$ parameters $\theta_1, \dots, \theta_k$, and let their unbiased least squares estimators $\hat{\theta}_1, \dots, \hat{\theta}_k$ be jointly normally distributed with $\text{var}(\hat{\theta}_i) = \sigma^2 \tau^2$ and $\text{corr}(\hat{\theta}_i, \hat{\theta}_j) = \rho$, where τ^2 and ρ are known constants depending on the design and σ^2 is the unknown error variance. Let S^2 be an unbiased estimator of σ^2 having ν df such that $\nu S^2/\sigma^2$ has a χ^2_ν distribution independent of the $\hat{\theta}_i$. Finally, let θ denote the vector $(\theta_1, \dots, \theta_k)$.

Typically the θ_i are contrasts among some population means. Three examples of this setting are: (1) orthogonal contrasts in a balanced design for which we have $\tau^2 = \sum_j c_{ij}^2$, where the c_{ij} are the contrast coefficients for the i th contrast θ_i ($1 \leq i \leq k$) and $\rho = 0$; (2) comparisons of k treatment means with a control mean in a one-way layout with n_0 observations on the control and n observations on each treatment, which leads to $\tau^2 = 1/n_0 + 1/n$ and $\rho = n/(n_0 + n)$; and (3) comparisons of k treatment means with a control mean in a balanced treatments incomplete block (BTIB) design, where the formulas for τ^2 and ρ are given by (4.3) and (4.4), respectively, of Bechhofer and Tamhane (1981).

2.2 Hypotheses and Test Procedures

Consider $k \geq 2$ upper one-sided hypothesis testing problems, $H_i : \theta_i = 0$ versus $K_i : \theta_i > 0$ ($1 \leq i \leq k$). Suppose that the type I FWE is to be controlled at a specified level α . In other words, if θ_m is any parameter configuration θ with $\theta_i = 0$, for $i = 1, \dots, m$ and $\theta_i > 0$ for $i = m+1, \dots, k$, then we require

$$P_{\theta_m}\{\text{Accept } H_1, \dots, H_m\} \geq 1 - \alpha,$$

$$m = 1, \dots, k. \quad (2.1)$$

Let $t_i = \hat{\theta}_i/s\tau$ (where s is the observed value of S) be the usual t statistic for testing H_i ($1 \leq i \leq k$). If H_1, \dots, H_m are true, then the corresponding random variables T_1, \dots, T_m have Student's m -variate central t distribution with ν df and associated common correlation coefficient ρ ($m = 1, \dots, k$). Let $c'_m = t_{m,\nu,\rho}^{(\alpha)}$ be the upper α point of $\max_{1 \leq i \leq m} T_i$, for $m = 1, \dots, k$. Bechhofer and Dunnett (1988) have given extensive tables of c'_m values. For $m = 1$, we have $c'_1 = t_\nu^{(\alpha)}$, the upper α point of Student's univariate t distribution with ν df.

The usual single-step test procedure (SS) rejects any H_i with $t_i \geq c'_k$, for example, Dunnett's (1955) procedure for the comparisons with a control problem. For the same problem, the usual step-down test procedure (SD), as proposed originally by Miller (1966, pp. 78 and 85–86) and later studied by Naik (1975) and Marcus, Peritz, and Gabriel (1976), first orders the statistics as $t_{(1)} \leq \dots \leq t_{(k)}$ (and the corresponding hypotheses as $H_{(1)}, \dots, H_{(k)}$), and then rejects any $H_{(j)}$ iff $H_{(j)}$ is rejected for $j = k, \dots, i + 1$ and $t_{(j)} \geq c'_i$. Both of these procedures control the type I FWE at level α ; see Hochberg and Tamhane (1987, chap. 2, Secs. 2.1.1 and 4.2). Since $c'_i < c'_k$ for $i < k$, it is clear that SD is uniformly more powerful than SS.

Our proposed step-up procedure (SU) will be based on another set of critical constants $c_1 \leq c_2 \leq \dots \leq c_k$ (to be defined shortly). It will accept $H_{(j)}$ iff $H_{(j)}$ is accepted for $j = 1, \dots, i - 1$ and $t_{(i)} < c_i$. Thus SU will begin by testing the smallest t statistic and work upward, accepting one hypothesis at a time and stopping by rejecting $H_{(i)}, H_{(i+1)}, \dots, H_{(k)}$ when $t_{(i)} \geq c_i$.

3. DETERMINATION OF CRITICAL CONSTANTS

3.1 Maximum Type I Familywise Error Rate

The critical constants of the SU procedure are to be determined so that the type I FWE requirement (2.1) is satisfied. Therefore, we first need to determine the maximum type I FWE or, equivalently, the minimum of the left side of (2.1), the minimum being taken over all θ_m , for each $m = 1, \dots, k$. This is done in the following theorem.

Theorem 3.1. Suppose that H_1, \dots, H_m are true and that H_{m+1}, \dots, H_k are false ($m = 1, \dots, k$). Then, for the SU procedure, the left side of (2.1) is minimized over θ_m when $\theta_i \rightarrow \infty$, for $i = m + 1, \dots, k$.

The proof of this theorem is given in Dunnett and Tamhane (1990) (hereafter abbreviated as DT), and is similar to the first part of the proof of Theorem 1 in Hayter and Tamhane (1990). In fact, using ideas similar to those in the second part of that proof, we can show that the left side of (2.1) is minimized over $\theta_i \leq 0$, for $i = 1, \dots, m$, when they are all equal to zero. In other words, we can consider the hypotheses $H_i : \theta_i \leq 0$ ($1 \leq i \leq k$) instead of $H_i : \theta_i = 0$ ($1 \leq i \leq k$) as considered here for convenience.

3.2 Properties of Critical Constants

To satisfy (2.1), the critical constants of the SU procedure must be determined so that the minimum found in Theorem 3.1 is at least $1 - \alpha$ for each $m = 1, 2, \dots$. The smallest possible value for each c_m is found by solving the following equation, obtained by setting this minimum equal to $1 - \alpha$:

$$P\{(T_1, \dots, T_m) < (c_1, \dots, c_m)\} = 1 - \alpha; \quad (3.1)$$

here T_1, \dots, T_m are as defined in Section 2.2, and $(x_1, \dots, x_m) < (y_1, \dots, y_m)$ denotes that $x_{1,m} < y_{1,m}, x_{2,m} < y_{2,m}, \dots, x_{m,m} < y_{m,m}$, where the $x_{i,m}$ and $y_{i,m}$ are the ordered x_i and y_i .

Although we have not been able to show analytically that solutions c_1, c_2, \dots, c_m satisfying the monotonicity condition $c_1 \leq c_2 \leq \dots \leq c_m < \infty$ exist for arbitrary m , we have been able to show this for $m = 2$; see DT. Moreover, our computational experience for $m \leq 8$ and $\alpha = .05$ indicates that they probably exist for the values of m likely to be encountered in practice. (In the case of independent test statistics, we have verified this for m as large as 1,000.) Finner, Roters, and Hayter (1991) have recently addressed this problem.

We next compare the critical constants used by the SU, SD, and Hochberg's (1988) step-up procedure (henceforth referred to as the HC procedure).

Proposition 3.1. Let c_m, c'_m , and $c''_m = t_{\nu}^{(\alpha/m)}$ ($m = 1, 2, \dots$) denote the critical constants used by the SU, SD, and HC procedures, respectively. We have

$$c_1 = c'_1 = c''_1 = t_{\nu}^{(\alpha)} \quad \text{and} \quad c'_m < c_m < c''_m, \quad m > 1. \quad (3.2)$$

Proof. The first part of (3.2) is obvious. The inequality $c'_m < c_m$ for $m > 1$ follows by comparing Equation (3.1) for c_m with the following equation for c'_m :

$$P\{T_{m,m} < c'_m\} = 1 - \alpha, \quad m = 1, \dots, k,$$

where $T_{m,m} = \max(T_1, \dots, T_m)$.

We have proved the inequality $c_m < c''_m$ only for $m = 2$ in DT, and we conjecture it to be true for $m > 2$.

Since the critical constants used by SU are larger than those used by SD, except for $m = 1$, when they are equal, we cannot say that SU will be uniformly more powerful than SD, as was the case between Hochberg's step-up procedure and Holm's step-down procedure. SU will, however, be uniformly more powerful than HC because of the inequality $c_m < c''_m$ for $m > 2$. This is, of course, not surprising since HC does not exploit the correlation structure between the test statistics as does SU.

As noted before, the critical constants c'_m are extensively tabulated by Bechhofer and Dunnett (1988). The critical constants c''_m , being the percentiles of the univariate Student's t distribution, can also be readily obtained. Exact computations show that c_m is much closer to c'_m than to c''_m ; see Table 5.

3.3 Computation of Critical Constants

From (3.1) we see that the critical constants must be computed recursively; to determine c_m , one must first know the values of c_1, \dots, c_{m-1} . This is in contrast to the computation of the critical constants $c'_m = t_{m,\nu,\rho}^{(\alpha)}$ for the SD procedure which, for different values of m , can be computed independently of each other. To compute c_m for $m > 1$, we now derive a general expression for the left side of (3.1) that can be evaluated efficiently on a computer.

Let Z_i , for $i = 0, 1, \dots, k$, be independent $N(0, 1)$ random variables, and let U be a $\sqrt{\chi^2_{\nu}/\nu}$ random variable independent of the Z_i . Then we can express $T_i = \{\sqrt{1 - \rho}Z_i - \sqrt{\rho}Z_0\}/U$ ($1 \leq i \leq m$) and write the desired probability as

$$\int_0^\infty \int_{-\infty}^\infty P\{(Z_1, \dots, Z_m) < (d_1, \dots, d_m)\} \phi(z_0) dz_0 f_\nu(u) du, \quad (3.3)$$

where $d_i = (c_i u + \sqrt{\rho z_0}) / \sqrt{1 - \rho}$ ($1 \leq i \leq m$), $\phi(z_0)$ is the standard normal density function, and $f_\nu(u)$ is the density function of U . A recursive formula for the probability term in the integrand of (3.3) is obtained in the following lemma.

Lemma 3.1. Let $\Phi(\cdot)$ denote the standard normal distribution function. We have

$$\begin{aligned} P\{(Z_1, \dots, Z_m) < (d_1, \dots, d_m)\} \\ &= P\{(Z_1, \dots, Z_{m-1}) < (d_2, \dots, d_m)\} \Phi(d_1) \\ &+ P\{(Z_1, \dots, Z_{m-1}) < (d_1, d_3, \dots, d_m)\} \{\Phi(d_2) - \Phi(d_1)\} \\ &\vdots \\ &+ P\{(Z_1, \dots, Z_{m-1}) < (d_1, \dots, d_{m-1})\} \{\Phi(d_m) - \Phi(d_{m-1})\}. \end{aligned} \quad (3.4)$$

Proof. The result follows by successively conditioning Z_m to lie in the intervals $(-\infty, d_1)$, $[d_1, d_2)$, \dots , $[d_{m-1}, d_m)$. Note that we do not condition Z_m to lie in the interval $[d_m, \infty)$ because, in that case, the event $\{(Z_1, \dots, Z_m) < (d_1, \dots, d_m)\}$ is not satisfied.

Let $\Phi_{12\dots m}$ denote the expression in (3.4). An algorithm for computing $\Phi_{12\dots m}$ is as follows.

Step 1. Calculate $\Phi_h = \Phi(d_h)$, for $h = 1, \dots, m$, a total of m terms.

Step 2. Calculate $\Phi_{hi} = \Phi_i \Phi_h + \Phi_h [\Phi_i - \Phi_h]$, for $1 \leq h < i \leq m$, a total of $\binom{m}{2}$ terms.

Step 3. Calculate $\Phi_{hij} = \Phi_{ij} \Phi_h + \Phi_{hj} [\Phi_i - \Phi_h] + \Phi_{hi} [\Phi_j - \Phi_i]$, for $1 \leq h < i < j \leq m$, a total of $\binom{m}{3}$ terms.

\vdots

Step m . Calculate $\Phi_{12\dots m} = \Phi_{2\dots m} \Phi_1 + \Phi_{13\dots m} [\Phi_2 - \Phi_1] + \dots + \Phi_{1\dots m-1} [\Phi_m - \Phi_{m-1}]$.

This algorithm was programmed and incorporated into a numerical integration routine to evaluate (3.3) and solve for the values of the constants c_m . The computational details can be found in DT. Due to the increasingly long computing times required as m increases, computations were restricted to $m \leq 8$. The values of the critical constants c_m for $\alpha = .05$, $df = 10, 20, 30$, and ∞ , $\rho = 0, .1, .2, .5$, and $m = 2(1)8$ are provided in Table 1 for one-sided tests discussed here, and in Table 2 for two-sided tests discussed in Section 6. For $m > 8$, we expect the c_m values to be very close to the c'_m values, and hence the latter may be used as good approximations (a bit on the liberal side, though).

4. POWER COMPARISONS WITH COMPETING PROCEDURES

4.1 Competing Procedures

The SU procedure was compared in terms of power with three other competing procedures, that meet the same type I FWE requirement (2.1). Two of these competing procedures were SS and SD, described in Section 2.2. The third one was Hochberg's (1988) step-up (HC) procedure, described in Section 1. As noted in Section 3.2, HC can be thought of as a conservative version of SU, and hence is uniformly less powerful than SU. It was included in the comparison to assess the loss in power due to the use of the conservative upper bounds c''_m rather than the exact critical constants c_m .

Two improvements of HC were also included in the power comparisons. One was by Rom (1990), who proposed a method [analogous to our equation (3.1)] for determining the exact critical values for HC; he also gave a recursive formula for computing these critical values, which are sharper than the conservative Bonferroni critical values c''_m used by

Table 1. Critical Constants c_m for the Step-Up Procedure (One-Sided Tests)*

ρ	df	$m = 1$	$m = 2$	$m = 3$	$m = 4$	$m = 5$	$m = 6$	$m = 7$	$m = 8$
0	10	1.813	2.220	2.443	2.600	2.722	2.821	2.904	2.977
	20	1.725	2.083	2.273	2.406	2.508	2.590	2.659	2.718
	30	1.697	2.040	2.221	2.347	2.443	2.520	2.584	2.639
	∞	1.645	1.960	2.123	2.235	2.319	2.386	2.442	2.490
.1	10	1.813	2.216	2.432	2.585	2.703	2.799	2.879	2.949
	20	1.725	2.080	2.266	2.396	2.496	2.576	2.644	2.701
	30	1.697	2.038	2.215	2.338	2.432	2.508	2.571	2.625
	∞	1.645	1.958	2.119	2.229	2.312	2.379	2.434	2.482
.3	10	1.813	2.201	2.401	2.541	2.649	2.736	2.810	2.873
	20	1.725	2.069	2.243	2.364	2.457	2.532	2.594	2.648
	30	1.697	2.028	2.194	2.310	2.398	2.469	2.528	2.579
	∞	1.645	1.950	2.102	2.207	2.286	2.350	2.403	2.448
.5	10	1.813	2.174	2.351	2.473	2.567	2.643	2.706	2.761
	20	1.725	2.047	2.203	2.310	2.392	2.458	2.513	2.561
	30	1.697	2.008	2.157	2.260	2.339	2.402	2.454	2.499
	∞	1.645	1.933	2.071	2.165	2.237	2.294	2.342	2.382

* $\alpha = .05$.

Table 2. Critical Constants c_m for the Step-Up Procedure (Two-Sided Tests)*

ρ	df	$m = 1$	$m = 2$	$m = 3$	$m = 4$	$m = 5$	$m = 6$	$m = 7$	$m = 8$
0	10	2.228	2.620	2.834	2.986	3.104	3.201	3.282	3.352
	20	2.086	2.419	2.597	2.723	2.820	2.898	2.964	3.020
	30	2.042	2.357	2.525	2.643	2.733	2.805	2.866	2.919
	∞	1.960	2.241	2.389	2.492	2.569	2.631	2.683	2.727
.1	10	2.228	2.619	2.832	2.983	3.100	3.195	3.275	3.345
	20	2.086	2.418	2.595	2.720	2.817	2.894	2.960	3.016
	30	2.042	2.356	2.523	2.640	2.730	2.802	2.863	2.915
	∞	1.960	2.241	2.388	2.490	2.567	2.629	2.681	2.725
.3	10	2.228	2.610	2.812	2.955	3.065	3.154	3.229	3.294
	20	2.086	2.410	2.581	2.700	2.791	2.865	2.927	2.980
	30	2.042	2.350	2.510	2.622	2.708	2.777	2.834	2.883
	∞	1.960	2.236	2.378	2.477	2.551	2.611	2.661	2.703
.5	10	2.228	2.587	2.770	2.897	2.994	3.073	3.139	3.196
	20	2.086	2.393	2.548	2.655	2.737	2.803	2.859	2.906
	30	2.042	2.334	2.481	2.582	2.659	2.721	2.773	2.818
	∞	1.960	2.223	2.355	2.445	2.513	2.568	2.614	2.653

* $\alpha = .05$.

HC. The other one by Hommel (1988) also uses the critical values c_m'' , but with a more complicated decision rule. Our results showed that the power increases for both of these procedures over HC were marginal at best, with Rom's procedure being slightly superior to Hommel's. Hence their results are not reported here.

4.2 Definition of Power

Let $\delta_i = \theta_i/\sigma\tau$ for $1 \leq i \leq k$, and let $\delta_m = (1/\sigma\tau)\theta_m$ denote the vector of the δ_i 's whose first m components are 0 and the last $k - m$ components are positive; that is, the first m hypotheses are true and the remaining $k - m$ are false. For any test procedure R based on the t_i -statistics, the power, as defined in (4.1), depends on the θ_i 's only through the δ_i 's. For $0 \leq m \leq k - 1$ and $1 \leq t \leq k - m$, we define the power of any procedure R under δ_m as

$$\Pi_{k,m,t}(R | \delta_m) = P_{\delta_m}\{R \text{ rejects at least } t \text{ of the } k - m \text{ false hypotheses}\}. \quad (4.1)$$

Our focus will be primarily on the $t = k - m$ case since, in most situations, one would be interested in detecting all false hypotheses; for instance, in drug development studies one would usually want to detect all the active drugs. To ensure that our comparisons are not biased by the choice of t , however, we will also consider the $t = 1$ case, which gives the probability of rejecting at least one false hypothesis.

4.3 Power Expressions

To derive expressions for (4.1) for various competing procedures R under consideration, we use the representation $T_i = \{\sqrt{1 - \rho}Z_i - \sqrt{\rho}Z_0\}/U$ ($1 \leq i \leq k$), where the Z_i are mutually independent normal random variables with unit variance, $E(Z_i) = 0$, for $i = 0, 1, \dots, m$, and $E(Z_i) = \delta_i/\sqrt{1 - \rho}$, for $i = m + 1, \dots, k$. For convenience, we will confine ourselves to the case $\delta_i = \delta$ for $i = m + 1, \dots, k$. It is easy to extend the results to the general case

where the δ_i are unequal; see DT. For the special case considered here, we denote the power expression in (4.1) simply by $\Pi_{k,m,t}(R | \delta)$ and the probability of an event \mathcal{E} computed under such δ_m by $P_{\delta_m}(\mathcal{E})$; we also put $\Delta = \delta/\sqrt{1 - \rho}$.

SU Procedure. For the SU procedure we get

$$\begin{aligned} \Pi_{k,m,t}(\text{SU} | \delta) &= \sum_{s=0}^{k-m-t} \sum_{r=0}^m \binom{k-m}{s} \binom{m}{r} \\ &\times P_{\delta_m}\{\text{accept } H_1, \dots, H_r, H_{m+1}, \dots, H_{m+s}; \\ &\text{reject } H_{r+1}, \dots, H_m, H_{m+s+1}, \dots, H_k\} \\ &= \sum_{s=0}^{k-m-t} \sum_{r=0}^m \binom{k-m}{s} \binom{m}{r} \\ &\times P_{\delta_m}\{(T_1, \dots, T_r, T_{m+1}, \dots, T_{m+s}) < (c_1, \dots, c_{r+s}); \\ &\min(T_{r+1}, \dots, T_m, T_{m+s+1}, \dots, T_k) \geq c_{r+s+1}\}. \quad (4.2) \end{aligned}$$

Now use the representation for the T_i 's defined in the preceding paragraph, condition on $U = u$ and $Z_0 = z_0$, and let $d_i = (c_i u + \sqrt{\rho} z_0)/\sqrt{1 - \rho}$ ($1 \leq i \leq k$). Then, using the independence of the Z_i 's ($1 \leq i \leq k$), unconditioning on U and Z_0 , and interchanging the order of summation and integration, (4.2) can be written as

$$\begin{aligned} \Pi_{k,m,t}(\text{SU} | \delta) &= \int_0^\infty \int_{-\infty}^\infty \sum_{s=0}^{k-m-t} \sum_{r=0}^m \binom{k-m}{s} \binom{m}{r} \\ &\times P\{(Z_1, \dots, Z_r, Z_{m+1}, \dots, Z_{m+s}) \\ &< (d_1, \dots, d_{r+s})\} [1 - \Phi(d_{r+s+1})]^{m-r} \\ &\times [1 - \Phi(d_{r+s+1} - \Delta)]^{k-m-s} \phi(z_0) dz_0 f_u(u) du. \quad (4.3) \end{aligned}$$

The probability term in (4.3) can be evaluated using a recursive formula similar to that derived in Lemma 3.1 after taking account of the fact that here the Z_i , for $i \geq m + 1$, have a nonzero mean Δ .

SD Procedure. For the SD procedure we get

$$\begin{aligned} \Pi_{k,m,t}(\text{SD} \mid \delta) &= \sum_{s=0}^{k-m-t} \sum_{r=0}^m \binom{k-m}{s} \binom{m}{r} \\ &\times P_{\delta,m}\{\text{accept } H_1, \dots, H_r, H_{m+1}, \dots, H_{m+s}; \\ &\text{reject } H_{r+1}, \dots, H_m, H_{m+s+1}, \dots, H_k\} \\ &= \sum_{s=0}^{k-m-t} \sum_{r=0}^m \binom{k-m}{s} \binom{m}{r} \\ &\times P_{\delta,m}\{\max(T_1, \dots, T_r, T_{m+1}, \dots, T_{m+s}) \\ &< c'_{r+s}, (T_{r+1}, \dots, T_m, T_{m+s+1}, \dots, T_k) \\ &\geq (c'_{r+s+1}, \dots, c'_k)\}, \end{aligned} \quad (4.4)$$

where $(x_1, \dots, x_m) \geq (y_1, \dots, y_m)$ denotes that $x_{1,m} \geq y_{1,m}$, \dots , $x_{m,m} \geq y_{m,m}$. Using the same method as that used in deriving (4.3), we can express (4.4) as

$$\begin{aligned} \Pi_{k,m,t}(\text{SD} \mid \delta) &= \int_0^\infty \int_{-\infty}^\infty \sum_{s=0}^{k-m-t} \sum_{r=0}^m \binom{k-m}{s} \binom{m}{r} \\ &\times P\{(Z_{r+1}, \dots, Z_m, Z_{m+s+1}, \dots, Z_k) \geq (d'_{r+s+1}, \dots, d'_k)\} \\ &\times \Phi'(d'_{r+s}) \Phi^s(d'_{r+s} - \Delta) \phi(z_0) dz_0 f_v(u) du, \end{aligned} \quad (4.5)$$

where the d'_i are defined analogous to the d_i . The probability term in (4.5) can be computed by using a recursive formula similar to that derived in Lemma 3.1 after taking account of the facts that the Z_i have a nonzero mean Δ , for $i \geq m+1$, and the direction of inequality is reversed here, making the formulas slightly different; the details can be found in DT.

SS Procedure. For the SS procedure we get

$$\begin{aligned} \Pi_{k,m,t}(\text{SS} \mid \delta) &= \sum_{s=0}^{k-m-t} \binom{k-m}{s} P_{\delta,m}\{T_i < c'_i, i = m+1, \dots, \\ &m+s; T_i \geq c'_i, i = m+s+1, \dots, k\} \\ &= \int_0^\infty \int_{-\infty}^\infty \sum_{s=0}^{k-m-t} \binom{k-m}{s} \{\Phi(d'_k - \Delta)\}^s \\ &\times \{1 - \Phi(d'_k - \Delta)\}^{k-m-s} \phi(z_0) dz_0 f_v(u) du. \end{aligned} \quad (4.6)$$

An alternative way of computing (4.6) would be to set $d'_i \equiv d'_k$, for $1 \leq i \leq k$, in (4.5).

HC Procedure. The power of the HC procedure is given by (4.3) with c_i replaced by $c''_i = t^{(\alpha/i)}$ ($1 \leq i \leq k$).

4.4 Power Results and Discussion

Proposition 4.1. When all k hypotheses are false (i.e., $m = 0$) and $t = k$, under any arbitrary configuration δ_0 (i.e., when the δ_i 's for $1 \leq i \leq k$ are nonzero and not necessarily equal), we have

$$\Pi_{k,0,k}(\text{SU} \mid \delta_0) = \Pi_{k,0,k}(\text{HC} \mid \delta_0) > \Pi_{k,0,k}(\text{SD} \mid \delta_0).$$

Proof. First note that $\Pi_{k,0,k}(\text{SU} \mid \delta_0) = P_{\delta_0}\{\min(T_1, \dots, T_k) \geq c_1\}$ and $\Pi_{k,0,k}(\text{HC} \mid \delta_0) = P_{\delta_0}\{\min(T_1, \dots, T_k) \geq$

$c''_1\}$, and the two are equal because $c_1 = c''_1$. Next, using $c_1 = c'_1$, we get

$$\begin{aligned} \Pi_{k,0,k}(\text{SU} \mid \delta_0) &= P_{\delta_0}\{\min(T_1, \dots, T_k) \geq c_1\} \\ &> P_{\delta_0}\{(T_1, \dots, T_k) \geq (c'_1, \dots, c'_k)\} \\ &= \Pi_{k,0,k}(\text{SD} \mid \delta_0), \end{aligned}$$

which completes the proof.

We conjecture that, if exactly one hypothesis is false and $t = 1$, then SD is uniformly more powerful than SU, that is, $\Pi_{k,k-1,1}(\text{SD} \mid \delta_{k-1}) > \Pi_{k,k-1,1}(\text{SU} \mid \delta_{k-1})$, for all δ_{k-1} . We have not been able to prove this conjecture, although numerical results support it (see Tables 3 and 5).

We studied the powers $\Pi_{k,m,t}$ of the four procedures, namely, SU, SD, SS, and HC, for $t = k - m, 1$, and for various values of $m = 0, 1, \dots, k - 1$ and $k = 2(1)6$. The results for $k = 5$ are shown in Table 3 when $k - m = 1, 3, 5$; the complete results are given in DT. First let us consider the results for $t = k - m$, which gives the probability of rejecting *all* false hypotheses. As shown in Proposition 4.1, we see that SU dominates SD uniformly when all hypotheses are false. On the other hand, when only one hypothesis is false, SD dominates SU uniformly. We further see that when an intermediate number of hypotheses are false, SU is more powerful than SD for small departures from the null values, while SD is more powerful in the other case. It is worth noting that the differences in power between SU and SD are quite small except when all (or most) hypotheses are false and the departures from the null values are not too large, in which case SU is more powerful than SD.

Turning to the comparison of SU with SS, we see that SS is more powerful than SU only when one hypothesis is false, and that, too, by a very small amount. In all other cases SU dominates SS by a wide margin. As noted in Section 3.2, HC is a conservative version of SU. It is much less powerful than SU when only a few hypotheses are false (but more powerful than SS except when exactly one hypothesis is false).

Next consider the power results for $t = 1$ and $k - m = 3.5$, which give the probabilities of rejecting *at least one* false hypothesis. Again SU dominates SD uniformly when all hypotheses are false. SD and SS have identical powers, which fact can be easily proved analytically. HC is the least powerful procedure using this definition of power, except when all hypotheses are false and δ is large.

The power comparison (using $t = k - m$) between SU and SD when either all hypotheses are false or exactly one hypothesis is false is made in more detail in Tables 4 and 5, respectively, for $k = 2(1)6$. In these tables δ values are so chosen as to make the power of SU equal to .50, .80, and .95. From Table 4 we see that, when all hypotheses are false, the advantage of SU increases with k and becomes relatively large particularly at low levels of power. On the other hand, from Table 5 we see that when exactly one hypothesis is false, the advantage of SD decreases with k and is in most cases negligibly small.

Table 3. Probability of Rejecting At Least t of the $k - m$ False Hypotheses [Power $\Pi_{k,m,t}(R | \delta)$] for Competing Procedures

Number of false H_i $k - m$	t	δ	Procedure			
			SU	SD	SS	HC
1	1	2	.408	.408	.408	.373
		3	.777	.778	.778	.750
		4	.961	.961	.961	.953
3	3	2	.228	.227	.173	.210
		3	.649	.650	.578	.625
		4	.933	.934	.906	.924
5	5	2	.293	.262	.106	.293
		3	.736	.715	.477	.736
		4	.963	.959	.866	.963
3	1	2	.659	.658	.658	.622
		3	.941	.940	.940	.928
		4	.997	.997	.997	.996
5	1	2	.775	.754	.754	.749
		3	.982	.971	.971	.979
		4	1.000	.999	.999	1.000

NOTE: $k = 5$, $\nu = \infty$, $\rho = .5$, and $\alpha = .05$.

5. p VALUES FOR MULTIPLE TEST PROCEDURES

For a multiple hypotheses testing problem, we define the p value for any hypothesis as the smallest overall significance level α at which that hypothesis can be rejected using a given multiple test procedure and the observed test statistics for all the hypotheses. We refer to such a p value as a *joint* p value, and, to distinguish from the so-called separate p value for each H_i that we used in Section 1, we will denote it by \bar{p} . Such joint (sometimes referred to as *adjusted*) p values have been used before by Westfall and Young (1989) among others. Once \bar{p} values are computed for a given procedure, they can be used with any fixed α to decide which hypotheses to reject.

We now give formulas for computing the \bar{p} values for the SS, SD, and SU procedures. In the following, T_1, \dots, T_m have an m -variate central t distribution with ν df and associated common correlation ρ . For the SS procedure we have

$$\bar{p}_{(m)} = P\{\max(T_1, \dots, T_k) \geq t_{(m)}\}, \quad \text{for } m = 1, \dots, k.$$

For the SD procedure, a formula for $\bar{p}_{(m)}$ was given in Dunnett and Tamhane (1991) (denoted there simply by p_m), which is as follows:

$$\begin{aligned} \bar{p}_{(m)} &= \bar{p}'_{(k)}, & \text{for } m = k, \\ &= \max(\bar{p}'_{(m)}, \bar{p}'_{(m+1)}), & \text{for } m = 1, \dots, k-1, \end{aligned}$$

where $\bar{p}'_{(m)}$ is given by

$$\bar{p}'_{(m)} = P\{\max(T_1, \dots, T_m) \geq t_{(m)}\}.$$

Table 4. Power Differences in Favor of SU Over SD, $\Pi_{k,m,k-m}(SU | \delta) - \Pi_{k,m,k-m}(SD | \delta)$, When All Hypotheses Are False ($m = 0$)

Power of SU	Number of H_i , k				
	2	3	4	5	6
.50	.013	.021	.027	.032	.036
.80	.007	.011	.014	.017	.020
.95	.002	.003	.004	.004	.005

NOTE: $\nu = \infty$, $\rho = .5$, and $\alpha = .05$.

For the SU procedure $\bar{p}_{(m)}$ is given by

$$\begin{aligned} \bar{p}_{(m)} &= \bar{p}'_{(1)}, & \text{for } m = 1, \\ &= \min(\bar{p}'_{(m)}, \bar{p}'_{(m-1)}), & \text{for } m = 2, \dots, k, \end{aligned}$$

where the computation of $\bar{p}'_{(m)}$ involves evaluating the constants $c_1 \leq \dots \leq c_m$ with $c_m = t_{(m)}$ such that the following set of equations is satisfied:

$$P\{(T_1, \dots, T_i) < (c_1, \dots, c_i)\} = 1 - \bar{p}'_{(m)}, \quad i = 1, \dots, m.$$

Note that this set of equations is similar to (3.1). In both cases c_1, \dots, c_{m-1} have to be evaluated, but here $c_m = t_{(m)}$ is given and $1 - \alpha = 1 - \bar{p}'_{(m)}$ has to be evaluated, while in (3.1), $1 - \alpha$ is given and c_m (in addition to c_1, \dots, c_{m-1}) has to be evaluated. The recursive formula derived in Lemma 3.1 can be used in this case, but the computations need to be done in an iterative manner starting with an initial guess at $\bar{p}'_{(m)}$, calculating c_1, \dots, c_{m-1} , and then finding a new value of $\bar{p}'_{(m)}$ from the last equation with $c_m = t_{(m)}$.

6. TWO-SIDED TESTS

All of the above theory can be extended in a straightforward manner to two-sided tests, where the alternatives to the $H_i : \theta_i = 0$ are now $K_i : \theta_i \neq 0$ ($1 \leq i \leq k$). The test statistics are $|t_i|$, and hence the determination of the critical constants and other analytical results involve the joint distribution of the random variables $|T_i|$. The critical constants for the SD procedure are given by $c'_m = |t_{(m), \nu, \rho}^{(\alpha)}|$, namely the upper α point of $\max_{1 \leq i \leq m} |T_i|$, for $m = 1, \dots, k$. The

Table 5. Power Differences in Favor of SD Over SU, $\Pi_{k,m,k-m}(SD | \delta) - \Pi_{k,m,k-m}(SU | \delta)$, When Exactly One Hypothesis Is False ($m = k - 1$)

Power of SU	Number of H_i , k				
	2	3	4	5	6
.50	.005	.003	.001	.000	.000
.80	.004	.002	.001	.000	.000
.95	.002	.001	.000	.000	.000

NOTE: $\nu = \infty$, $\rho = .5$, and $\alpha = .05$.

Table 6. Critical Constants for One-Sided Tests for Different Procedures Used in Examples

Procedure	Constant	<i>m</i>					
		1	2	3	4	5	6
SU	c_m	1.697	2.008	2.157	2.260	2.339	2.402
SD	c_m	1.697	1.989	2.147	2.255	2.335	2.399
SS	c'_k	2.399	2.399	2.399	2.399	2.399	2.399
HC	c''_m	1.697	2.042	2.231	2.360	2.457	2.536

NOTE: $\nu = 30$, $\rho = 0.5$, and $\alpha = 0.05$.

common critical constant used by SS is $c'_k = |t_{k,\nu,\rho}^{(\alpha)}|$. These constants have been tabulated by Bechhofer and Dunnett (1988). For the two-sided SU procedure, the critical constants c_m satisfy the relation [analogous to (3.1)]:

$$P\{|T_1|, \dots, |T_m| < (c_1, \dots, c_m)\} = 1 - \alpha,$$

where (T_1, \dots, T_m) have a joint m -variate central t distribution with ν df and a common associated correlation coefficient ρ . The critical constants c_m for the two-sided case are given in Table 2.

7. EXAMPLES

We now give two examples to illustrate how the results of the procedures considered here compare on the same data. Suppose $k = 6$, $\rho = .5$, $\nu = 30$, and one-sided tests are to be done at an overall significance level $\alpha = .05$. The critical constants used by different procedures are listed in Table 6. The data for the two examples, namely the t_m statistics (labeled according to their ordered values) and the corresponding separate p_m values, are listed in Table 7. The joint \tilde{p}_m values for the multiple test procedures SS, SD, and SU are also listed in this table (as well as the intermediate calculations of \tilde{p}'_m values for SD and SU). Note that, because the t_m are assumed to be already ordered, that is $m = (m)$, we do not use the (m) notation in the following discussion and in Tables 6 and 7.

Example 1. First we see that SS rejects only H_6 since only $t_6 \geq c'_k = 2.399$. Next we see that SD rejects H_6 but accepts H_5 and hence, by implication, also all the remaining hypotheses. (This can be deduced by noting that $t_5 = 2.320 < c'_5 = 2.335$ or by noting that $\tilde{p}_m > .05$ for $m \leq 5$.) SU

accepts H_1 but rejects H_2 and hence, by implication, also all the remaining hypotheses. (This can be deduced by noting that $t_2 = 2.020 > c_2 = 2.008$ or by noting that $\tilde{p}_m \leq .05$ for $m \geq 2$.) HC can be applied in a step-up manner either by comparing the ordered p_m values with α/m or, equivalently, by comparing the ordered t_m statistics with the critical constants c''_m ; it accepts H_1 and H_2 but rejects H_3 , and hence, by implication, also all the remaining hypotheses. Thus, in this example, SU rejects most hypotheses (H_2 through H_6) followed by HC; both SD and SS reject only H_6 .

Example 2. First we see that SS rejects H_5 and H_6 and accepts the rest. Next we see that SD rejects H_2 through H_6 and only accepts H_1 . SU accepts H_1 through H_3 but rejects H_4 , and hence, by implication, also all the remaining hypotheses. HC accepts H_1 through H_4 but rejects H_5 and hence, by implication, also H_6 . Thus, in this example, SD rejects most hypotheses, followed by SU, followed by HC and SS.

8. EXTENSIONS AND CONCLUDING REMARKS

Although in this article we have confined our attention to the case of equicorrelated and equivariance estimates $\hat{\theta}_i$, it would be useful to extend SU to the case where $\text{var}(\hat{\theta}_i) = \sigma^2 \tau_i^2$ and $\text{corr}(\hat{\theta}_i, \hat{\theta}_j) = \rho_{ij}$; here the τ_i^2 and ρ_{ij} are known constants depending on the design. An important special case is the problem of comparing treatments with a control in an unbalanced one-way lay-out, for which $\tau_i^2 = 1/n_0 + 1/n_i$ and $\rho_{ij} = \lambda_i \lambda_j$, where $\lambda_i = \sqrt{n_i/(n_0 + n_i)}$, n_0 and n_i denoting the sample sizes for the control and the i th treatment, respectively ($1 \leq i \leq k$). The proof of Theorem 3.1 makes crucial use of the equal correlation as-

Table 7. Statistics and p Values for Two Examples

Example	Procedure	Statistic and p value	<i>m</i>					
			1	2	3	4	5	6
1	SU	t_m	1.50	2.02	2.25	2.28	2.32	2.50
		p_m	.072	.026	.016	.015	.014	.009
		\tilde{p}'_m	.072	.049	.041	.048	.052	.041
		\tilde{p}_m	.072	.049	.041	.041	.041	.041
		\tilde{p}'_m	.072	.047	.041	.048	.052	.041
	SD	\tilde{p}_m	.072	.052	.052	.052	.052	.041
		\tilde{p}_m	.245	.105	.068	.064	.059	.041
	SS	\tilde{p}_m	.245	.105	.068	.064	.059	.041
		\tilde{p}_m	.245	.105	.068	.064	.059	.041
		\tilde{p}_m	.245	.105	.068	.064	.059	.041
2	SU	t_m	1.50	2.00	2.15	2.30	2.47	2.50
		p_m	.072	.027	.020	.014	.010	.009
		\tilde{p}'_m	.072	.051	.051	.046	.038	.041
		\tilde{p}_m	.072	.051	.051	.046	.038	.038
		\tilde{p}'_m	.072	.049	.050	.046	.038	.041
	SD	\tilde{p}_m	.072	.050	.050	.046	.041	.041
		\tilde{p}_m	.245	.109	.082	.061	.043	.041
	SS	\tilde{p}_m	.245	.109	.082	.061	.043	.041
		\tilde{p}_m	.245	.109	.082	.061	.043	.041
		\tilde{p}_m	.245	.109	.082	.061	.043	.041

sumption. If the theorem is true in the unequal correlation case, then a different proof must be found. The computing of the c_i for this case would have to be done specifically for each problem. The probability expression to be computed would be as in Equation (3.3) but with $d_i = (c_i + \lambda_{iz_0})/\sqrt{1 - \lambda_i^2}$ ($1 \leq i \leq k$). Alternatively, approximate values of the c_i could be computed by using the arithmetic average of the p_{ij} between the θ_i still left to be tested. Whether the resulting approximation is conservative would require investigation.

Finally we note that, whereas the uniform improvement in power that was found to occur when the step-up modification of the Bonferroni procedure due to Hochberg (1988) was used instead of the step-down modification due to Holm (1979) may have suggested that step-up procedures have some inherent power advantage, we have found from our comparison of the two approaches under normal theory that the step-up testing has a nonnegligible power advantage only in those situations where most hypotheses are false and it is desired to reject all of them. This power advantage increases with the number of false hypotheses. But our comparison under normal theory also has revealed that even in situations where only a few hypotheses are false, the step-up test procedure stands at only a negligible power disadvantage with respect to the step-down test procedure.

[Received May 1990. Revised December 1990.]

REFERENCES

- Bechhofer, R. E., and Dunnett, C. W. (1988), "Tables of Percentage Points of Multivariate Student t Distributions," in *Selected Tables in Mathematical Statistics*, 11, Providence, RI: American Mathematical Society, pp. 1-371.
- Bechhofer, R. E., and Tamhane, A. C. (1981), "Incomplete Block Designs for Comparing Treatments With a Control: General Theory," *Technometrics*, 23, 45-57.
- Carmer, S. G., and Walker, W. M. (1982), "Baby Bear's Dilemma: A Statistical Tale," *Agronomy Journal*, 74, 122-124.
- Duncan, D. B. (1965), "A Bayesian Approach to Multiple Comparisons," *Technometrics*, 7, 171-222.
- Duncan, D. B., and Dixon, D. O. (1983), " k -ratio t Tests, t Intervals, and Point Estimates for Multiple Comparisons," in *Encyclopedia of Statistics* (vol. 4) eds. S. Kotz and N. L. Johnson, New York: John Wiley, pp. 403-410.
- Dunnett, C. W. (1955), "A Multiple Comparison Procedure for Comparing Several Treatments With a Control," *Journal of the American Statistical Association*, 50, 1096-1121.
- Dunnett, C. W., and Tamhane, A. C. (1990), "A Step-up Multiple Test Procedure," Technical Report 90-1, Northwestern University, Dept. of Statistics.
- (1991), "Step-Down Multiple Tests for Comparing Treatments With a Control in Unbalanced One-way Layouts," *Statistics in Medicine*, 10, 939-947.
- Finner, H., Roters, M., and Hayter, A. J. (1991), "On the joint distribution function of order statistics with reference to step-up multiple test procedures," manuscript submitted for publication.
- Hayter, A. J., and Tamhane, A. C. (1990), "Sample Size Determination for Step-Down Multiple Test Procedures: Orthogonal Contrasts and Comparisons With a Control," *Journal of Statistical Planning and Inference*, 27, 271-290.
- Hochberg, Y. (1988), "A Sharper Bonferroni Procedure for Multiple Tests of Significance," *Biometrika*, 75, 800-802.
- Hochberg, Y., and Tamhane, A. C. (1987), *Multiple Comparison Procedures*, New York: John Wiley.
- Holm, S. (1979), "A Simple Sequentially Rejective Multiple Test Procedure," *Scandinavian Journal of Statistics*, 6, 65-70.
- Hommel, G. (1988), "A Stagewise Rejective Multiple Test Procedure Based on a Modified Bonferroni Test," *Biometrika*, 75, 383-386.
- Marcus, R., Peritz, E., and Gabriel, K. R. (1976), "On Closed Testing Procedures With Special Reference to Ordered Analysis of Variance," *Biometrika*, 63, 655-660.
- Miller, R. G., Jr. (1966), *Simultaneous Statistical Inference*, New York: McGraw-Hill.
- Naik, U. D. (1975), "Some Selection Rules for Comparing p Processes With a Standard," *Communications in Statistics, Part A—Theory and Methods*, 4, 519-535.
- O'Brien, P. C. (1983), "The Appropriateness of Analysis of Variance and Multiple Comparison Procedures," *Biometrics*, 39, 787-788.
- O'Neill, R. T., and Wetherill, G. B. (1971), "The Present State of Multiple Comparisons Methods" (with discussion), *Journal of Royal Statistical Society, Ser. B*, 33, 218-241.
- Perry, J. N. (1986), "Multiple-Comparison Procedures: A Dissenting View," *Journal of Economic Entomology*, 79, 1149-1155.
- Rom, D. M. (1990), "A Sequentially Rejective Test Procedure Based on a Modified Bonferroni Inequality," *Biometrika*, 77, 663-665.
- Rothman, K. J. (1990), "No Adjustments Are Needed for Multiple Comparisons," *Epidemiology*, 1, 43-46.
- Scheffé, H. (1953), "A Method of Judging All Contrasts in the Analysis of Variance," *Biometrika*, 40, 87-104.
- Tukey, J. W. (1953), *The Problem of Multiple Comparisons*, unpublished manuscript, Princeton University, Dept. of Statistics.
- Waller, R. A., and Duncan, D. B. (1969), "A Bayes Rule for Symmetric Multiple Comparisons," *Journal of the American Statistical Association*, 80, 1484-1503.
- Welsch, R. E. (1977), "Stepwise Multiple Comparison Procedures," *Journal of the American Statistical Association*, 72, 566-575.
- Westfall, P. H., and Young, S. S. (1989), " p Value Adjustments for Multiple Tests in Multivariate Binomial Models," *Journal of the American Statistical Association*, 84, 780-786.