BIOMETRICS 55, 1258–1262
December 1999

Rank-Based Tests for Dose Finding in Nonmonotonic Dose–Response Settings

Yuh-Ing Chen

Institute of Statistics, National Central University Chung-Li, Taiwan 32054
email: ychen@stat.ncu.edu.tw

SUMMARY. Lim and Wolfe (1997, *Biometrics* **53,** 410–418) proposed rank-based multiple test procedures for identifying the dose levels that are more effective than the zero-dose control in randomized complete block designs when it can be assumed that the efficacy of the increasing dose levels is monotonically increasing up to a point, followed by a monotonic decrease. Modifications of the Lim–Wolfe tests are suggested that provide more practical and powerful alternatives. Two numerical examples are illustrated and the results of a Monte Carlo power study are presented.

KEY WORDS: Dose–response study; Lim–Wolfe test; Monte Carlo study; Randomized block design; Step-down closed test; Umbrella-pattern treatment effects.

1. Introduction

In a dose–response study for drug development, several increasing dose levels of a certain drug are usually compared with the zero-dose control to assess the effect of the drug. When experiment units are subject to relatively large variability, a randomized complete block design is often conducted where experiment units are divided into more homogeneous blocks and the experiment units in each block are randomly assigned to receive the doses under study. If the experimenter has the prior information that increasing dose levels would produce stronger or at least equal treatment effects, House (1986) proposed a nonparametric testing procedure to find the lowest dose level for which the response is stochastically larger than that at the zero-dose control. However, it occurs frequently that the dose–response relationship is anticipated to follow an umbrella pattern, monotonically increasing up to a point, followed by a monotonic decrease. The point separating the dose–response relationship into two different orderings is referred to as the peak of the umbrella (Mack and Wolfe, 1981). Lim and Wolfe (1997) suggested multiple testing procedures to determine the dose levels that are more effective than the zero-dose control for the umbrella-pattern dose–response relationship. They considered two cases. The peak-known test requires knowledge of the point where the dose–response pattern changes from increasing to decreasing. Their peak-unknown test has a cumbersome feature of requiring the estimation of the unknown peak at each stage of the procedure. In this paper, we consider more convenient and practical alternatives to the Lim–Wolfe testing procedures.

In Section 2, the House and Lim–Wolfe tests are briefly described. In Section 3, the Lim–Wolfe tests are shown to be closed tests in the sense of Marcus, Peritz, and Gabriel (1976). Modifications of the Lim–Wolfe tests are then proposed for both cases where the peak of the umbrella is known or unknown. In Section 4, two numerical examples demonstrate the feasibility of the proposed tests. Finally, Section 5 reports the results of a Monte Carlo simulation investigation of the relative power performances of the competing tests for a variety of umbrella-patterned treatment effects configurations.

2. The House and Lim–Wolfe Tests

Let Y_{ij} for $i = 1, \ldots, n$, $j = 0, 1, \ldots, k$ be independent continuous random variables with the distribution function of Y_{ij} given by $F(x - \beta_i - \theta_j)$, where the β_i's are block effects that are not of direct interest and the θ_j's are treatment effects. Suppose that the zero treatment ($i = 0$) is the zero-dose control and the other k treatments correspond to increasing dose levels. House (1986) discusses the problem of contrasting increasing dose levels of a substance in a randomized complete block design. Let $\bar{R}_{j:0}, \bar{R}_{j:1}, \ldots, \bar{R}_{j:j}$ be Friedman's (1937) average ranks obtained from the previous $j+1$ treatments and set $\hat{R}_1^{(j)} \leq \cdots \leq \hat{R}_j^{(j)}$ to be the isotonic regression estimators of $\bar{R}_{j:1}, \ldots, \bar{R}_{j:j}$ under the ordered restriction $\theta_1 \leq \cdots \leq \theta_j$. Define

$$T_j = \left(\hat{R}_j^{(j)} - \bar{R}_{j:0}\right) \{V_j(2/n)\}^{-1/2}, \qquad j = 1, \ldots, k, \quad (2.1)$$

where $V_j = (j+1)(j+2)/12$. Let $\bar{t}_\alpha(n, j)$ be the upper αth percentile of T_j. House (1986) then suggests, at the first step, claiming $\theta_k > \theta_0$ if $T_k \geq \bar{t}_\alpha(n, k)$; otherwise, stop the procedure and claim $\theta_j = \theta_0$, $j = 1, \ldots, k$. If the test based on T_k rejects, then proceed to claim $\theta_{k-1} > \theta_0$ if $T_{k-1} \geq \bar{t}_\alpha(n, k-1)$. Continue the procedure until it stops. If the test stops at, say, the $(k - j^* + 2)$th step where the first $j^* - 1$ dose levels are comparing with the zero-dose control, House then concludes that j^* is the lowest dose level such that $\theta_{j^*} > \theta_0$. Note that, if ties occur within a block for the first j treat-

ments and the control, the average ranks are used and the factor V_j is reduced by H_j, where any group of t tied ranks contributes $(t^3 - t)/(12jn)$ to H_j.

Lim and Wolfe (1997) consider identification of the dose levels that are more effective than the zero-dose control under umbrella-patterned treatment effects in a randomized complete block design. When the umbrella peak p among k groups is known *a priori*, they first find the isotonic regression estimators $\hat{R}^{(p)}_{k:1} \le \cdots \le \hat{R}^{(p)}_{k:p} \ge \cdots \ge \hat{R}^{(p)}_{k:k}$ of the Friedman's (1937) average ranks $\bar{R}^{(p)}_{k:1}, \ldots, \bar{R}^{(p)}_{k:k}$ under the umbrella-pattern restriction $\theta_1 \le \cdots \le \theta_p \ge \cdots \ge \theta_k$. Let $t_\alpha(n, k; p)$ be the upper αth percentiles of the null distribution of

$$T_{k:p} = \left(\hat{R}^{(p)}_{k:p} - \bar{R}_{k:0}\right) \{V_k(2/n)\}^{-1/2}. \tag{2.2}$$

Lim and Wolfe then suggest claiming $\theta_p > \theta_0$ if $T_{k:p} \ge t_\alpha(n, k; p)$; otherwise, stop the procedure and claim $\theta_j = \theta_0$, $j = 1, \ldots, k$. If the test based on $T_{k:p}$ rejects, then they delete the pth treatment and apply the Friedman's ranking scheme on the remaining observations in the $k - 1$ treatments and the control to obtain the average ranks. Assuming that the peak of the umbrella is currently known at u among the $k-1$ groups, they further find the isotonic regression estimators under the appropriate umbrella-pattern restriction with peak at u. After computing the statistic $T_{k-1:u}$ in the same way, Lim and Wolfe claim $\theta_u > \theta_0$ if $T_{k-1:u} \ge t_\alpha(n, k-1; u)$. Continue the procedure until it stops. Suppose that s and r are the lowest and highest dose levels that are significantly better than the control. Lim and Wolfe then conclude that $\theta_j > \theta_0$, $j = s, \ldots, r$. Note that, at a certain stage where k' ($<k$) treatments relabeled from 1 to k' with known peak at q (excluding the significant treatment levels) are compared with the control, the test statistic would be

$$T_{k':q} = \left(\hat{R}^{(q)}_{k':q} - \bar{R}_{k':0}\right) \{V_{k'}(2/n)\}^{-1/2}. \tag{2.3}$$

Also note that, when $p = k$, the Lim–Wolfe peak-known test is identical to House's (1986) test. Moreover, to use the test, experimenters must specify clearly the umbrella peak at each stage of the procedure, which, however, does not occur very often in practical situations.

For the case with unknown umbrella peak, Lim and Wolfe (1997) estimate the unknown peak as $\hat{p}$, which satisfies

$$Q_{\hat{p}} = \min\left\{ Q_t = \sum_{j=1}^{k'} \left(\bar{R}_{k':j} - \hat{R}^{(t)}_{k':j}\right)^2, \qquad t = 1, \ldots, k' \right\}$$

when there are k' treatments remaining to be compared with the control. They then suggest implementing the testing scheme for the peak-known setting based on the statistics

$$\tilde{T}_{k':\hat{p}} = \left(\hat{R}^{(\hat{p})}_{k':\hat{p}} - \bar{R}_{k':0}\right) \{V_{k'}(2/n)\}^{-1/2}, \tag{2.4}$$

comparing with the critical value $\tilde{t}_\alpha(n, k')$, the upper αth percentile of the null distribution of $\tilde{T}_{k':\hat{p}}$. Note that, if there are treatments tied for having the minimum Q_t, let χ be the set of groups tied for the minimum Q_t. Lim and Wolfe (1997) then take the value of $\tilde{T}_{k':\hat{p}}$ to be the average of the $T_{k':j}$'s for those j in the set χ. In addition, to utilize the Lim–Wolfe peak-unknown test, experimenters need to estimate the unknown peak based on the statistics Q_t at each stage of the procedure, which requires considerable computational effort.

3. Modifications of the Lim–Wolfe Procedures

The problem of determining the dose levels that are more effective than the zero-dose control in an umbrella-pattern dose–response relationship can be formulated as a sequence of hypothesis testing problems as follows:

$$\mathrm{H}_0^{(u,v)} : (\theta_0 = \theta_1 = \cdots = \theta_u = \theta_v = \cdots = \theta_k)$$

versus

$$\mathrm{H}_1^{(u,v)} : (\theta_0 \le \theta_1 \le \cdots \le \theta_{u-1} \le \theta_u \quad \text{and} \quad \theta_v \ge \theta_{v+1} \ge \cdots \ge \theta_k \ge \theta_0) \tag{3.1}$$

for $u \le v = 1, 2, \ldots, k$. If u^* and v^* are the smallest u and largest v, respectively, for which $\mathrm{H}_0^{(u,v)}$ is rejected, then we claim $\theta_j > \theta_0$, $j = u^*, \ldots, v^*$.

Note that the family of null hypotheses $\mathrm{H}_0 = \{\mathrm{H}_0^{(u,v)}\}$ is closed under intersection in the sense that $\mathrm{H}_0^{(u,v)} \in \mathrm{H}_0$ and $\mathrm{H}_0^{(u',v')} \in \mathrm{H}_0$ imply $\mathrm{H}_0^{(u,v)} \cap \mathrm{H}_0^{(u',v')} \in \mathrm{H}_0$. According to Marcus et al. (1976), a level-α closed procedure that includes separate level-α tests of individual $\mathrm{H}_0^{(u,v)}$ applied in a step-down manner can be employed to determine the dose levels that are more effective than the zero-dose control. Note that the closed testing scheme strongly controls the familywise error rate (FWE), which is the probability that at least one true $\mathrm{H}_0^{(u,v)}$ is rejected. Also note that the Lim–Wolfe (1997) procedures starting with $u = v = p$ or $\hat{p}$ are closed testing procedures for H_0. In this section, however, we propose different closed testing procedures for comparing umbrella-pattern treatment effects with a control in a randomized complete block design.

When the peak of the umbrella is known as p, we claim at the first step $\theta_p > \theta_0$ if $T_{k:p} \ge t_\alpha(n, k; p)$; otherwise, stop the procedure and claim $\theta_j = \theta_0$, $j = 1, \ldots, k$. However, if the test based on $T_{k:p}$ rejects, we suggest using House's (1986) procedure separately for comparing the two different groups of treatments $\{1, \ldots, p-1\}$ and $\{p+1, \ldots, k\}$ with the control for identifying the lowest and highest dose levels that are more effective than the zero-dose control. Note that the modified Lim–Wolfe peak-known test is equivalent to House's test when $p = k$ and can be regarded as an extension of Chen's (1993) test to randomized complete block designs.

For the unknown-peak setting, we suggest estimating the unknown umbrella peak among the k treatments based on the statistics Q_t, say $\hat{p}$, and test for $\mathrm{H}_0^{(\hat{p},\hat{p})}$ based on $\tilde{T}_{k:\hat{p}}$. If $\tilde{T}_{k:\hat{p}} \ge \tilde{t}_\alpha(n, k)$, then proceed using House's (1986) multiple procedure for comparing treatments $\{1, \ldots, \hat{p}-1\}$ and $\{\hat{p}+1, \ldots, k\}$, respectively, with the control to identify the dose levels that are more effective than the zero-dose control.

To appreciate why the Lim–Wolfe (1997) tests need to be modified, note that the modified Lim–Wolfe peak-known test specifies only the peak of the umbrella among the k treatments, which is of more practical use than the original Lim–Wolfe peak-known test. Moreover, the modified Lim–Wolfe peak-unknown test estimates only the unknown umbrella peak among the k treatments at the first step, saving considerable computational work, especially when k is large.

4. Two Examples

4.1 *The Lim–Wolfe Example*

Consider first the data set analyzed in Lim and Wolfe (1997) that corresponds to an experimental design discussed in Heffner, Drawbaugh, and Zigmond (1974) in studying the effect of the drug d-amphetamine sulfate on the behavior of rats. Ten male albino rats of the same strain and of approximately the same weight were utilized. Five dose levels of the drug, specified in terms of milligrams of drug per kilogram of weight of the rat, were studied, namely 0.0, 0.5, 1.0, 1.5, and 2.0 mg/kg, where the zero level consists of a saline solution. One hour after a drug dosage injection was administered, an experimental session began during which the rat received water each time after a second lever was pressed. Each rat received all five dose levels in a random order. The observation is the lever press rate defined as the total number of lever presses divided by the elapsed time (in seconds) during a session for a given dosage.

To implement the modified Lim–Wolfe peak-unknown procedure, first find Friedman's average ranks $\bar{R}_{4:0} = 1.45$, $\bar{R}_{4:1} = 2.85$, $\bar{R}_{4:2} = 4.35$, $\bar{R}_{4:3} = 4.0$, and $\bar{R}_{4:4} = 2.35$ and compute $Q_1 = 1.2317$, $Q_2 = 0.0$, $Q_3 = 0.0613$, and $Q_4 = 2.2817$. The estimated umbrella peak is $\hat{p} = 2$. The correction for ties is

$$3(2^3 - 2)/(12 \cdot 10 \cdot 4) = 0.0375$$

and the statistic

$$\tilde{T}_{4:2} = (4.35 - 1.45)\{(5 \cdot 6/12 - 0.0375)(2/10)\}^{-1/2} = 4.132$$

is greater than $\tilde{t}_{0.05}(10, 4) = 2.121$ obtained from Table 1 in Lim and Wolfe (1997). So we claim, at the 5% level, that the 1.0 mg/kg dosage results a higher lever press rate than that of 0.0 mg/kg, a saline solution.

Now, for comparing the dosages 0.0 and 0.5 mg/kg, we obtain $\bar{R}_{1:0} = 1.20$ and $\bar{R}_{1:1} = 1.80$. The statistic $\mathrm{H}_1 = (1.80 - 1.20)/\{(2 \cdot 3/12)(2/10)\}^{1/2} = 1.897$ is above 1.645. Hence, there is an effect at 0.5 mg/kg.

Next, applying House's (1986) procedure to the dosages 0.0, 2.0 (treatment 1), and 1.5 (treatment 2) mg/kg, we have $\bar{R}_{2:0} = 1.25$, $\bar{R}_{2:1} = 1.85$, and $\bar{R}_{2:1} = 2.90$, respectively. The correction for ties is

$$(2^3 - 2)/(12 \cdot 10 \cdot 2) = 0.025$$

and the statistic

$$\mathrm{H}_2 = (2.90 - 1.25)\{(3 \cdot 4/12 - 0.025)(2/10)\}^{-1/2} = 3.337$$

is larger than $\bar{t}_{0.05}(10, 2)$ or $\bar{t}_{0.05}(10, 2; 2) = 1.677$, obtained from Table 1 in Lim and Wolfe (1997). So there is an effect at 1.5 mg/kg.

Finally, comparing the dosage 2.0 mg/kg with the zero-dose control, the average ranks are $\bar{R}_{1:0} = 1.25$ and $\bar{R}_{1:1} = 1.75$. The correction for ties is

$$(2^3 - 2)/(12 \cdot 10 \cdot 1) = 0.05$$

and the statistic

$$\mathrm{H}_1 = (1.75 - 1.25)\{(2 \cdot 3/12 - 0.05)(2/10)\}^{-1/2} = 1.667$$

is higher than 1.645. Therefore, the modified Lim–Wolfe procedure reaches the same conclusion as stated in Lim and Wolfe (1997) that, at the level-$\alpha = 0.05$, there is an effect at all four dose levels. However, the modified Lim–Wolfe peak-unknown test is more convenient to implement since it does not have to estimate the unknown umbrella peak at each stage of the procedure as required in the original Lim–Wolfe procedure.

Table 1
Ranks of observations for three dose level comparison

	Block									
	1	2	3	4	5	6	7	8	9	10
Control	1	1	2	2	1	1	2	2	2	2
Dose 1	2	3	1	3	3	3	1	3	1	1
Dose 2	4	4	4	4	4	4	4	4	4	4
Dose 3	3	2	3	1	2	2	3	1	3	3

4.2 *An Artificial Example*

As a second example, consider the comparison of three dose levels with a control in a randomized complete block design with 10 blocks in which the observations have the ranks as shown in Table 1.

The average ranks are obtained as $\bar{R}_{3:0} = 1.6$, $\bar{R}_{3:1} = 2.1$, $\bar{R}_{3:2} = 4$, and $\bar{R}_{3:3} = 2.3$. The values of $Q_1 = 1.805$, $Q_2 = 0.0$, and $Q_3 = 1.445$ yield an estimated umbrella peak $\hat{p} = 2$. The statistic given by

$$\tilde{T}_{4:2} = (4 - 1.6)\{(4 \cdot 5/12)(2/10)\}^{-1/2} = 4.157$$

is greater than $\tilde{t}_{0.01}(10.3) = 2.598$ obtained from Table 2 in Lim and Wolfe (1997). So there is very strong evidence of an effect at dose level 2.

Now, examining dose level 1, the rank averages $\bar{R}_{1:0} = 1.4$ and $\bar{R}_{1:1} = 1.6$ give

$$\mathrm{H}_1 = (1.6 - 1.4)/\{(2 \cdot 3/12)(2/10)\}^{1/2} = 0.632,$$

which is even less than 1.282, the critical value for $\alpha = 0.10$. So there is a nonsignificant effect at dose level 1.

Examining dose level 3, we obtain the average ranks $\bar{R}_{1:0} = 1.2$ and $\bar{R}_{1:1} = 1.8$, giving

$$\mathrm{H}_1 = (1.8 - 1.2)/\{(2 \cdot 3/12)(2/10)\}^{1/2} = 1.897.$$

For $\alpha = 0.05$, the critical value is 1.645. So at the 5% level, there is evidence of an effect at dose level 3. Therefore, we conclude that, at the 5% level, there is evidence of an effect at both dose levels 2 and 3.

Note the original Lim–Wolfe (1997) peak-unknown test also claims that there is very strong evidence of an effect at dose level 2 at the first step. At the second step, they would compare the dose levels 1 and 3 (relabled as 2) with the control. The average ranks $\bar{R}_{2:0} = 1.6$, $\bar{R}_{2:1} = 2.1$, and $\bar{R}_{2:2} = 2.3$ give $Q_1 = 0.02$ and $Q_2 = 0.0$. The estimated umbrella peak is $\hat{p} = 2$. Having found that the statistic

$$\tilde{T}_{2:2} = (2.3 - 1.6)\{(3 \cdot 4/12)(2/10)\}^{-1/2} = 1.565$$

is less than $\tilde{t}_{0.05}(10, 2) = 2.012$, Lim and Wolfe would not investigate dose level 1, and they should conclude that there is evidence of an effect only at dose level 2.

Table 2
Estimated powers for $k = 4$ and $n = 10$ at $\alpha = 0.05$

θ_{10}	θ_{20}	θ_{30}	θ_{40}		LW(p)	MLW(p)	LW($\hat{p}$)	MLW($\hat{p}$)
Normal Distribution								
0	0	0.5	1	π_{30}	0.244	0.244	0.146	0.170
				π_{40}	0.571	0.571	0.437	0.500
0	0.5	1	0	π_{20}	0.246	0.257	0.149	0.211
				π_{30}	0.602	0.602	0.443	0.578
0	0.5	1	1.5	π_{20}	0.203	0.203	0.149	0.179
				π_{30}	0.610	0.610	0.443	0.503
				π_{40}	0.867	0.867	0.786	0.812
0.5	1	1.5	0	π_{10}	0.233	0.246	0.162	0.186
				π_{20}	0.610	0.648	0.456	0.490
				π_{30}	0.880	0.880	0.791	0.816
Exponential Distribution								
0	0	0.5	1	π_{30}	0.387	0.387	0.248	0.255
				π_{40}	0.778	0.778	0.657	0.669
0	0.5	1	0	π_{20}	0.394	0.411	0.251	0.261
				π_{30}	0.772	0.772	0.663	0.681
0	0.5	1	1.5	π_{20}	0.356	0.356	0.265	0.277
				π_{30}	0.811	0.811	0.679	0.687
				π_{40}	0.961	0.961	0.923	0.936
0.5	1	1.5	0	π_{10}	0.360	0.366	0.270	0.278
				π_{20}	0.806	0.846	0.675	0.680
				π_{30}	0.959	0.959	0.920	0.927

5. Monte Carlo Study

We conducted a Monte Carlo study to examine the relative power performances of the original Lim–Wolfe tests, LW(p) and LW($\hat{p}$), and the modified Lim–Wolfe tests, MLW(p) and MLW($\hat{p}$), for identifying the dose levels that are more effective than the zero-dose control in an umbrella-pattern dose–response relationship with known or unknown umbrella peak in a randomized block design. The study was performed for comparing $k = 4$ treatments with a zero-dose control, with $n = 10$ blocks in each case, and for a variety of patterns of treatment effects.

For each of these settings, appropriate normal and exponential deviates were derived by the IMSL routines RNNOR and RNEXP, respectively. The normal distributions under consideration have the unit variance but different means, and the exponential distributions have various location parameters with the common unit scale parameter. The designated alternative configurations correspond to values of $\theta_{i0} = \theta_i - \theta_0$, $i = 1, 2, \ldots, k$. The pairwise powers (the probability of declaring the ith treatment better than the control), denoted by π_{i0}, are simulated. In each case, we used 10,000 replications in obtaining the various power estimates. The simulated pairwise power estimates for the four tests are presented in Table 2.

We observe from Table 2 that the powers of the peak-known tests LW(p) and MLW(p) are the same for comparing ordered treatment effects with the control. This is not surprising since both tests are equivalent to House's (1986) test for the monotonic dose–response relationship. Moreover, for comparing umbrella-pattern treatment effects with the control, the tests LW(p) and MLW(p) have the same power for the peak group-control comparison. For the detection of effects at the remaining treatments excluding the peak group, however, the MLW(p) test is more powerful than the LW(p) test. The simulation results also indicate that the power of the MLW($\hat{p}$) test is, in general, higher than that of the LW($\hat{p}$) test for comparing umbrella-pattern treatment effects with the control.

In conclusion, the use of the modified Lim–Wolfe (1997) tests are recommended for identifying the dose levels that are more effective than the zero-dose control in randomized block designs for two reasons. First, the modified peak-known test is of more practical use and the modified peak-unknown test is more convenient to implement than the corresponding original Lim–Wolfe tests. Second, the modified tests have better power performances than their respective original tests.

Acknowledgements

The author wishes to thank the editor, associate editor, and two referees for their useful suggestions and comments, which lead to an improved presentation of the results.

Résumé

Lim et Wolfe (1997, *Biometrics* **53,** 410–418) ont proposé des procédures de tests multiples des rangs afin d'identifier les niveaux de doses les plus efficaces par rapport à la dose zéro de contrôle dans des dispositifs en blocs complets randomisés lorsque l'on peut considérer que l'efficacité des doses croissantes augmente de façon monotone jusqu'à un certain niveau, suivi d'une décroissance monotone. Des modifications des tests de Lim-Wolfe (1997) sont proposées, qui fournissent des solutions plus pratiques et plus puissantes. Deux exemples numériques illustrent la méthode et les résultats d'une étude de puissance par des simulations de Monte-Carlo sont présentés.

References

Chen, Y. I. (1993). Nonparametric comparisons of umbrella pattern treatment effects with a control in a one-way layout. *Communications in Statistics—Simulation and Computation* **22,** 749–764.

Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association* **32,** 675–701.

Heffner, T. G., Drawbaugh, R. B., and Zigmond, M. J. (1974). Amphetamine and operant behavior in rats: Relationship between drug effect and control response rate. *Journal of Comparative and Physiological Psychology* **86,** 1031–1043.

House, D. E. (1986). A nonparametric version of Williams' test for a randomized block design. *Biometrics* **42,** 187–190.

Lim, D. H. and Wolfe, D. A. (1997). Nonparametric tests for comparing umbrella pattern treatment effects with a control in a randomized block design. *Biometrics* **53,** 410–418.

Mack, G. A. and Wolfe, D. A. (1981). K-sample rank tests for umbrella alternatives. *Journal of the American Statistical Association* **76,** 175–181.

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.

Received November 1998. *Revised February* 1999. *Accepted March* 1999.