An Approximate Likelihood Ratio Test for Comparing Several Treatments to a Control

Dei-In Tang; Shang P. Lin

Journal of the American Statistical Association, Vol. 92, No. 439. (Sep., 1997), pp. 1155-1162.

Stable URL:
http://links.jstor.org/sici?sici=0162-1459%28199709%2992%3A439%3C1155%3AAALRTF%3E2.0.CO%3B2-G

Journal of the American Statistical Association is currently published by American Statistical Association.

An Approximate Likelihood Ratio Test for Comparing Several Treatments to a Control

Dei-In TANG and Shang P. LIN

Some parametric tests for comparing several treatments to a control are examined. Those tests developed earlier tend to rest on a single principle: simplicity or power; those developed later attempt to fill the gap. We propose a new test that represents another such attempt. The basic idea is to find a simple statistic that is a good approximation to the likelihood ratio statistic. This leads to a power function similar to that of the likelihood ratio test, which is optimal among the available competitors. Like the orthogonal contrast test, the new test is built on an orthogonal relationship, resulting in a relatively simple null distribution that depends on the sample sizes only through their total. Thus critical values can be easily computed and a detailed listing results in only a moderately large table. Computing the statistic is elementary when the sample sizes of the new treatments are equal; otherwise, some matrix operations are needed. A simple procedure is given to estimate the sample size required to achieve a given power level. To compete with Dunnett's test in terms of pairwise comparison, the new test (or any of the other tests) can be applied according to the closed testing procedure. In this setting the new test is compared to Dunnett's test by simulation.

KEY WORDS: Closed testing; Cone hypothesis; Dunnett's test; Multiple comparisons; Multivariate one-sided testing; Orthant hypothesis.

1. INTRODUCTION

Many experiments, particularly in drug development, test more than one new treatment in the hope of finding at least one that is better than the control in terms of the mean of a response variable. Thus a natural strategy in the statistical analysis is to compare each new treatment individually to the control based on a test of significance. This entails the problem of multiple testing, to which the standard solution is to adjust the level of each test to control the overall level; that is, the experimentwise type I error rate. In this vein Dunnett (1955) gave a parametric method, and Steel (1959) gave a nonparametric method.

But the natural strategy may not be very powerful. This is because the correlation structure is not utilized in constructing the test statistic, although the adjustment made for multiple testing has to take this structure into account. An alternative strategy is to treat the problem by one multivariate test based on all the data. Several authors (Akkerboom 1990; Bartholomew 1961; Conaway, Pillers, Robertson, and Sconing 1991; Mukerjee, Robertson, and Wright 1987) adopted this strategy and considered the same set of null and alternative hypotheses. The null hypothesis states that all new treatments are as good as the control; the alternative states that all new treatments are as good as or better than the control and at least one is better. In the same framework, we propose a new test.

Note that the alternative hypothesis is one-sided because inequalities are used. The relationship among the mean parameters defined by these inequalities is known as a simple tree order. It is a common view that considering a one-sided alternative leads to more powerful tests. But caution should be taken to interpret the result from such a test. In particular, without a priori knowledge that strongly supports the assumption of one-sidedness, it may be misleading to interpret the rejection of the null hypothesis as evidence supporting the alternative hypothesis, especially when a multivariate test is used. In earlier work (Tang and Lin 1994) we presented a more detailed discussion and a treatment of this concern in a similar context.

Having given some specifics, it is useful to reconsider the difference between the two strategies. Because the natural strategy considers more than one set of null and alternative hypotheses, it can give a more detailed conclusion, which can include statements such as that treatment 1 is better than the control. In other words, tests under the first strategy are also applicable under the second strategy, but not vice versa. However this advantage can be realized only if there is sufficient power to conclude existence of better treatments. On the other hand, approaches that attempt only to conclude existence of better treatments without identifying any of them seems unsatisfactory in the context considered. Thus a powerful multivariate test needs a supplemental procedure to identify better treatments. Fortunately, such procedures have been developed; for example, the closed testing procedure of Marcus, Peritz, and Gabriel (1976). Note that the idea of closed testing and another idea can also improve on Dunnett's test, but with respect to power (Dunnett and Tamhane 1991, 1992).

Some tests in the multivariate framework are examined. All of them assume normally distributed data, and most of them were motivated by geometric insights. The region of the alternative hypothesis is a polyhedral cone; that of the null hypothesis is a linear subspace contained in the boundary of the cone. Bartholomew (1961) developed the likelihood ratio (LR) test. The power of this test is generally good, but computing the critical value is a challenging task that has generated a significant amount of research. Now there is a detailed listing of critical values in the case in which all sample sizes are equal. In the case of unequal sample sizes, there are approximation techniques as well as computer programs. Power approximations are also avail-

Dei-In Tang is Research Scientist and Shang P. Lin is Research Scientist, Statistical Sciences and Epidemiology Division, Nathan S. Kline Institute for Psychiatric Research, Orangeburg, NY 10962. This research was supported in part by National Institutes of Health grants MH42959 and CA54708.

Journal of the American Statistical Association
September 1997, Vol. 92, No. 439, Theory and Methods

able (see Robertson, Wright, and Dykstra 1988 for details of this body of knowledge and relevant references). Although the development of the LR test seems quite complete, many practitioners might find the amount of details excessive. An alternative strategy to make the LR test work is to develop a test with a similar, but simpler power function. This can be done by simplifying the LR statistic in the way of approximation. To this end, it is useful to view the LR statistic as a maximum contrast, with every direction in the cone of the alternative hypothesis providing a contrast (see Mukerjee et al. 1987, for a detailed explanation). The simplest approximation uses only one direction, resulting in a t test. Effort is required to secure an optimal direction. Such is the approach of Abelson and Tukey (1963). The same test was derived by Schaafsma and Smid (1966) based on a different optimality criterion. But the power of a single-contrast test can be very low in some part of the cone, because the tree-order cone is "wide." Hence the test is inappropriate for general use. Mukerjee et al. (1987) studied a class of multiple-contrast tests that includes Dunnett's test and, as an extreme, the aforementioned single-contrast test. They recommended the orthogonal contrast (OC) test for being simple and nearly as powerful as the LR test. A drawback is the assumption of equal sample size for all new treatments. Another—and perhaps more important—drawback is that the power tends to dip near the center of the cone, as illustrated later in Table 2. This has an undesirable design implication, to be described.

Conaway et al. (1991) and, independently, Akkerboom (1990) used a circular cone to approximate the original cone and developed the corresponding LR test. Intuitively, as long as the new cone has a large overlap with the original cone, the new LR test should closely approximate the original LR test in power. The new LR test is simpler, because the new cone is geometrically simpler. The idea of using the LR test corresponding to a simpler cone to replace the original more difficult LR test was also discussed by Tang, Gnecco, and Geller (1989), who in some sense considered a more general problem. We follow their approach in deriving our test. Instead of a circular cone, we use an orthant (cone). To secure a large overlap, we align the two center directions, suitably defined. In contrast, the circular cone of Conaway et al. (1991) requires that two things be determined: the center direction and the angle. Determining an appropriate angle takes effort but can result in a better approximation than can be obtained by an orthant. An unfortunate consequence is that the null distribution depends on the angle, making tabulation of critical values difficult. Both the circular-cone and orthant tests improve on the OC test by allowing arbitrary sample sizes. The distributional simplicity of the orthant test has facilitated the development of some group sequential procedures in the case of known variances (Tang, Geller, and Pocock 1993), although this advantage is beyond the scope of our discussion.

In Section 2 we describe the new test. In Section 3 we provide simulation results of this test for comparison with results given by Conaway et al. (1991). In Section 4 we address two design issues in using the new test: how to allocate a given total of subjects among the treatments and control and what is the total to achieve a given power. In Section 5 we simulate the performance of the new test, the LR test, and Dunnett's test in the closed testing setting. We give concluding remarks in Section 6.

2. AN APPROXIMATE LR TEST

Suppose that there are k new treatments, labeled 1, 2, $\ldots, k$. Label the control treatment with 0. Let $\{x_{ij}, j = 1, \ldots, n_i\}$ be a sample for treatment i, $i = 0, \ldots, k$. Denote the sample mean $n_i^{-1} \sum_j x_{ij}$ by y_i and the total sample size $\sum_{0 \le i \le k} n_i$ by N. Assume that the samples are independent and follow $N(\theta_i, \sigma^2)$. Consider the problem of testing

$$H\colon \theta_i = \theta_0 \qquad (i = 1, \ldots, k) \quad \text{versus} \quad K\colon \theta_i \ge \theta_0$$
$$(i = 1, \ldots, k) \quad \text{with at least one inequality strict.} \quad (1)$$

We treat only the case of unknown σ^2. Results for known σ^2, corresponding to $N = \infty$, are given as Remark 1 later.

In terms of $\mu_i = \theta_i - \theta_0$, the problem becomes testing

$$H'\colon \mu_i = 0 \qquad (i = 1, \ldots, k) \quad \text{versus} \quad K'\colon \mu_i \ge 0$$
$$(i = 1, \ldots, k) \quad \text{with at least one inequality strict.} \quad (2)$$

These are the hypotheses considered by Tang et al. (1989). But transforming the data into a form amenable to their approach is not as straightforward.

Derivation of the LR statistic is standard (see Robertson et al. 1988, p. 63). The maximum likelihood estimate (MLE) of the common value of the θ_i's under H is $N^{-1} \sum_{0 \le i \le k} n_i y_i$, which we denote by $\bar{y}$. Let θ_i^* denote the MLE of θ_i under $K \cup H$. These θ's solve the problem min $\sum_{0 \le i \le k} n_i (y_i - \theta_i)^2$ subject to $\theta_j \ge \theta_0$, $j = 1, \ldots, k$. A simple algorithm to compute θ_i^* was described by Robertson et al. (1988, ex. 1.3.2). The LR test rejects H if $\lambda_{\text{LR}} > c$, where

$$\lambda_{\text{LR}} = \sum_{0 \le i \le k} n_i (\theta_i^* - \bar{y})^2$$
$$\div \left(\sum_{0 \le i \le k} \sum_j (x_{ij} - y_i)^2 + \sum_{0 \le i \le k} n_i (y_i - \bar{y})^2 \right).$$

(The constant c may be found in table A.8 of Robertson et al. 1988 if $n_0 = \cdots = n_k$.)

To describe the new statistic, we put λ_{LR} in an alternative form. Let $\mathbf{z}' = (z_1, \ldots, z_k)$, where $z_i = y_i - y_0$. The covariance matrix of $\mathbf{z}$ is $\sigma^2 \mathbf{\Omega}$, where the ij element of $\mathbf{\Omega}$ is $n_i^{-1} + n_0^{-1}$ for $i = j$ and n_0^{-1} otherwise.

Lemma. For any two $(k+1)$ vectors $\mathbf{a} = (a_0, \ldots, a_k)'$ and $\mathbf{b} = (b_0, \ldots, b_k)'$, define an inner product $(\mathbf{a}, \mathbf{b})_n = \sum n_i a_i b_i$. Suppose that $\sum n_i a_i = \sum n_i b_i = 0$. Let $\mathbf{c} = (a_1 - a_0, \ldots, a_k - a_0)'$ and $\mathbf{d} = (b_1 - b_0, \ldots, b_k - b_0)'$. Then $(\mathbf{a}, \mathbf{b})_n = \mathbf{c}' \mathbf{\Omega}^{-1} \mathbf{d}$.

The proof is omitted. This lemma implies $\sum n_i(y_i - \bar{y})^2 = \mathbf{z}'\boldsymbol{\Omega}^{-1}\mathbf{z}$ and $\sum n_i(y_i - \theta_i^*)^2 = (\mathbf{z} - \boldsymbol{\mu}^*)'\boldsymbol{\Omega}^{-1}(\mathbf{z} - \boldsymbol{\mu}^*)$, where $\boldsymbol{\mu}^*$ denotes the solution to the problem $\min(\mathbf{z} - \boldsymbol{\mu})'\boldsymbol{\Omega}^{-1}(\mathbf{z} - \boldsymbol{\mu})$ subject to $\mu_i \geq 0$, $i = 1, \ldots, k$. The two identities imply that $\sum_{0 \leq i \leq k} n_i(\theta_i^* - \bar{y})^2 = \boldsymbol{\mu}^{*\prime}\boldsymbol{\Omega}^{-1}\boldsymbol{\mu}^*$. Thus

$$\lambda_{\mathrm{LR}} = \boldsymbol{\mu}^{*\prime}\boldsymbol{\Omega}^{-1}\boldsymbol{\mu}^*/(S_q + \mathbf{z}'\boldsymbol{\Omega}^{-1}\mathbf{z}), \tag{3}$$

where S_q denotes $\sum_{0 \leq i \leq k} \sum_j (x_{ij} - y_i)^2$. Note that $\boldsymbol{\mu}^{*\prime}\boldsymbol{\Omega}^{-1}\boldsymbol{\mu}^*$ is the LR statistic for testing the hypotheses in (2) based on $\mathbf{z}$, if σ is known to be 1. The approach of Tang et al. (1989) that replaces this statistic by an approximation can now be followed. Let $\mathbf{J}$ denote the vector of k 1's, $\mathbf{e}_i$ denote the ith unit coordinate vector of k dimensions, $\mathbf{D}$ denote the diagonal matrix with diagonal elements $(\mathbf{e}_i'\boldsymbol{\Omega}^{-1}\mathbf{e}_i)^{-1/2}$, $i = 1, \ldots, k$. Let $\mathbf{A}$ be a square matrix satisfying that

$$\text{(a)}\ \mathbf{A}'\mathbf{A} = \boldsymbol{\Omega}^{-1}$$

and

$$\text{(b)}\ \mathbf{J}'\mathbf{A}\mathbf{D} = p\mathbf{J}' \quad \text{for a} \quad p > 0. \tag{4}$$

The purpose of (b) is explained in Remark 3 later. Let $\hat{\boldsymbol{\mu}}$ denote the solution to the problem $\min(\mathbf{z} - \boldsymbol{\mu})'\boldsymbol{\Omega}^{-1}(\mathbf{z} - \boldsymbol{\mu})$ subject to $\mathbf{A}\boldsymbol{\mu} \geq 0$. Following Tang et al. (1989), we replace $\boldsymbol{\mu}^{*\prime}\boldsymbol{\Omega}^{-1}\boldsymbol{\mu}^*$ in (3) by $\hat{\boldsymbol{\mu}}'\boldsymbol{\Omega}^{-1}\hat{\boldsymbol{\mu}}$, yielding

$$\lambda_{\mathrm{ALR}} = \hat{\boldsymbol{\mu}}'\boldsymbol{\Omega}^{-1}\hat{\boldsymbol{\mu}}/(S_q + \mathbf{z}'\boldsymbol{\Omega}^{-1}\mathbf{z}).$$

There are infinitely many choices of $\mathbf{A}$ that satisfy (4), each leading to a separate test. The Appendix describes the procedure of Tang et al. (1989) for obtaining a specific choice, which requires matrix operations. For the common special case when $n_1 = \cdots = n_k$, the symmetry of $\boldsymbol{\Omega}$ leads to the following simpler choice:

$$\mathbf{A} = \boldsymbol{\Omega}^{-1/2} = n^{1/2}(I - k^{-1}\{1 - [n_0/(n_0 + kn)]^{1/2}\}\mathbf{J}\mathbf{J}'),$$

where n denotes the common sample size. With $\mathbf{A}$ found, $\hat{\boldsymbol{\mu}}'\boldsymbol{\Omega}^{-1}\hat{\boldsymbol{\mu}}$ is computed as $\sum(w_i^+)^2$, where $\mathbf{w} = \mathbf{A}\mathbf{z}$ and w_i^+ denotes $\max(0, w_i)$. Computing λ_{ALR} is then easy.

To derive the null distribution, put $\lambda_{\mathrm{ALR}} = \sum_{1 \leq i \leq k} (w_i^+)^2/(S_q + \sum_{1 \leq i \leq k}(w_i^+)^2 + \sum_{1 \leq i \leq k}(w_i^-)^2)$, where w_i^- denotes $\max(0, -w_i)$. Because the w_i's are iid $N(0, \sigma^2)$ and $\sigma^{-2}S_q$ is independent of $\mathbf{w}$ and distributed as $\chi^2(N - k - 1)$, a chi-squared variable with $N - k - 1$ df, we have

$$P\{\lambda_{\mathrm{ALR}} \geq c|H\} = \sum_{0 \leq i \leq k} \pi_i P\{B_{i/2,(N-i-1)/2} \geq c\}, \tag{5}$$

where $\pi_i = 2^{-k}k!/(i!(k - i)!)$ and $B_{a,b}$ denotes a beta variable with parameters a and b. Here we assume $N > k + 1$ and define $B_{0,b}$ as the constant 0. The foregoing distribution belongs to the family of $\bar{E}^2$ distributions (Robertson et al. 1988). Although λ_{LR} has a similar null distribution, the associated weights are much more difficult to evaluate (see theorem 2.3.1 of Robertson et al. 1988). Critical values for $\lambda_{\mathrm{ALR}'}$ computed in GAUSS (Aptech Systems, Inc. 1991), are listed in Table 1.

Remark 1. For the case of known σ^2, say $\sigma^2 = 1$, the LR test can be based on the numerator of λ_{LR}. The corresponding ALR statistic is the numerator of λ_{ALR}. To obtain the null distribution of this ALR statistic, replace $B_{i/2,(N-i-1)/2}$ in (5) by $\chi^2(i)$. Corresponding critical values appear in Table 1 as rows with $N - k - 1 = \infty$.

Remark 2. The ALR test and the OC test have a close connection. The orthogonal contrasts of the OC test define an orthant that, through the same process, transforms into the orthant of the ALR test. Note that the OC test assumes the condition $n_1 = \cdots = n_k$. Thus, the simpler choice of $\mathbf{A}$ is used here. As a result, the OC statistic can be computed as $\max(w_i)/(S_q + \mathbf{z}'\boldsymbol{\Omega}^{-1}\mathbf{z})^{1/2}$.

Remark 3. The condition (b) in (4) aligns the center direction of the approximation cone with that of the original cone. As a result, in our simulation experience the ALR test shows greatest power in the center of the original cone, as the LR test does. Here, "center direction" means the direction in the cone that is equi-angled with all of the edge directions; however, in the case of the tree-order cone, the edge directions are specially defined (see Mukerjee et al. 1987). The tendency of the power of the OC test to dip in the center of the cone may be explained by the lack of the center as a contrast. The tendency becomes more obvious with increasing dimensionality because the angle between the center and any of the orthogonal contrasts grows bigger.

Remark 4. The problem considered in this article is invariant both under a scale change and under a permutation of the new treatments. Usually, one wants the test to be invariant. The single-contrast, LR, and Dunnett tests all qualify. The OC test qualifies but it requires the condition $n_1 = \cdots = n_k$, under which both the circular cone and ALR tests also qualify. However, without this condition the last two tests may not be permutation invariant. We give an easy remedy but argue that the defect is not serious. Tang et al. (1989) have shown how to make the ALR test invariant with more computation. Here we give a much easier scheme that works for both tests in question. Assume for brevity that all treatment sample sizes are distinct. The scheme, as part of the test procedure, is to order the treatments, no matter how they are initially ordered, according to decreasing sample size, or any other ordering of the sample sizes. With the original procedure, the order of the treatments is given a priori, not as part of the procedure. Thus the invariance issue here seems superficial. What is important is to recognize that the ordering of the treatments can affect the test statistic, and thus replication of analysis, if ever needed, should be based on the same ordering. This mainly conceptual nuisance seems bearable in the pursuit of a simple approximation.

3. POWER COMPARISON

The relative power of the ALR test as compared to the other tests may be anticipated in light of the foregoing discussion. Here we present actual power values to confirm our understanding as well as to show the size of the dif-

Table 1. Critical Values of the $\bar{E}^2$ Distribution (C = Entry Value * $(k/2)/(N - k/2 - 1)$)

	α = .05								
$N-k-1 \backslash k$	2	3	4	5	6	7	8	9	10
1	1.942	1.596	1.411	1.293	1.209	1.146	1.097	1.056	1.023
2	2.553	1.995	1.700	1.517	1.390	1.297	1.226	1.169	1.122
3	2.923	2.269	1.914	1.690	1.535	1.422	1.334	1.265	1.208
4	3.163	2.465	2.077	1.828	1.654	1.526	1.427	1.348	1.283
5	3.330	2.613	2.205	1.940	1.753	1.614	1.506	1.420	1.350
6	3.452	2.728	2.308	2.032	1.837	1.690	1.575	1.484	1.408
7	3.546	2.820	2.393	2.110	1.908	1.755	1.636	1.540	1.460
8	3.619	2.894	2.464	2.177	1.969	1.813	1.689	1.590	1.507
9	3.679	2.956	2.524	2.234	2.023	1.863	1.737	1.634	1.550
10	3.728	3.009	2.576	2.283	2.070	1.908	1.779	1.675	1.588
11	3.769	3.054	2.621	2.327	2.112	1.948	1.818	1.711	1.623
12	3.803	3.092	2.660	2.365	2.150	1.984	1.852	1.745	1.655
13	3.833	3.126	2.695	2.400	2.183	2.017	1.884	1.775	1.684
14	3.859	3.156	2.725	2.430	2.214	2.046	1.913	1.803	1.711
15	3.882	3.182	2.753	2.458	2.241	2.073	1.939	1.828	1.736
16	3.902	3.206	2.778	2.483	2.266	2.098	1.963	1.852	1.759
17	3.920	3.227	2.800	2.506	2.289	2.121	1.985	1.874	1.780
18	3.937	3.246	2.821	2.527	2.310	2.141	2.006	1.894	1.800
19	3.951	3.264	2.839	2.546	2.329	2.161	2.025	1.913	1.819
20	3.964	3.279	2.856	2.564	2.347	2.179	2.043	1.931	1.836
21	3.976	3.294	2.872	2.580	2.364	2.195	2.060	1.948	1.853
22	3.987	3.307	2.887	2.596	2.379	2.211	2.075	1.963	1.868
23	3.997	3.320	2.900	2.610	2.394	2.226	2.090	1.978	1.883
24	4.007	3.331	2.913	2.623	2.407	2.239	2.104	1.991	1.896
25	4.015	3.341	2.924	2.635	2.420	2.252	2.117	2.004	1.909
26	4.023	3.351	2.935	2.646	2.432	2.264	2.129	2.016	1.921
27	4.030	3.360	2.945	2.657	2.443	2.275	2.140	2.028	1.933
28	4.037	3.369	2.955	2.667	2.453	2.286	2.151	2.039	1.944
29	4.044	3.377	2.964	2.677	2.463	2.296	2.161	2.049	1.954
30	4.050	3.385	2.972	2.686	2.472	2.306	2.171	2.059	1.964
40	4.093	3.440	3.035	2.753	2.544	2.379	2.246	2.136	2.042
50	4.120	3.475	3.075	2.797	2.590	2.427	2.296	2.187	2.094
60	4.138	3.499	3.102	2.827	2.622	2.461	2.331	2.223	2.132
80	4.161	3.529	3.137	2.865	2.663	2.506	2.378	2.272	2.182
100	4.175	3.547	3.159	2.890	2.690	2.534	2.407	2.303	2.214
120	4.184	3.560	3.173	2.906	2.707	2.553	2.428	2.324	2.236
160	4.195	3.575	3.192	2.927	2.730	2.577	2.454	2.352	2.265
200	4.202	3.585	3.203	2.939	2.744	2.592	2.470	2.369	2.283
300	4.212	3.597	3.218	2.957	2.763	2.613	2.492	2.392	2.308
∞*	4.231	3.623	3.249	2.992	2.802	2.656	2.538	2.441	2.360
					α = .01				
1	1.997	1.662	1.491	1.384	1.308	1.251	1.205	1.166	1.134
2	2.897	2.245	1.912	1.707	1.567	1.465	1.385	1.322	1.270
3	3.596	2.711	2.256	1.976	1.786	1.648	1.542	1.459	1.391
4	4.124	3.081	2.537	2.201	1.973	1.807	1.680	1.580	1.498
5	4.529	3.376	2.769	2.391	2.133	1.944	1.800	1.687	1.595
6	4.845	3.617	2.962	2.553	2.271	2.065	1.907	1.783	1.681
7	5.098	3.816	3.126	2.692	2.392	2.172	2.003	1.868	1.759
8	5.305	3.983	3.267	2.813	2.498	2.266	2.088	1.946	1.830
9	5.477	4.125	3.388	2.919	2.592	2.351	2.164	2.016	1.895
10	5.622	4.247	3.493	3.012	2.676	2.426	2.234	2.080	1.954
11	5.745	4.353	3.586	3.095	2.751	2.495	2.297	2.138	2.008
12	5.851	4.446	3.669	3.169	2.818	2.557	2.354	2.191	2.058
13	5.944	4.527	3.742	3.236	2.879	2.613	2.406	2.241	2.104
14	6.026	4.600	3.808	3.296	2.935	2.665	2.455	2.286	2.147
15	6.099	4.665	3.867	3.351	2.986	2.712	2.499	2.328	2.187
16	6.163	4.724	3.921	3.400	3.032	2.756	2.540	2.367	2.223
17	6.221	4.777	3.970	3.446	3.075	2.796	2.578	2.403	2.258
18	6.273	4.825	4.015	3.488	3.114	2.834	2.614	2.437	2.290
19	6.321	4.869	4.056	3.527	3.151	2.868	2.647	2.468	2.320
20	6.364	4.910	4.094	3.562	3.185	2.901	2.678	2.498	2.349
21	6.404	4.947	4.129	3.596	3.216	2.931	2.707	2.525	2.375
22	6.440	4.981	4.161	3.627	3.246	2.959	2.734	2.551	2.400
23	6.474	5.013	4.191	3.655	3.273	2.985	2.759	2.576	2.424
24	6.504	5.042	4.219	3.682	3.299	3.010	2.783	2.599	2.446
25	6.533	5.070	4.246	3.707	3.324	3.034	2.806	2.621	2.468
26	6.560	5.095	4.270	3.731	3.346	3.056	2.827	2.642	2.488

(continued)

Table 1 (continued)

$N-k-1\backslash k$	$\alpha = .01$								
	2	3	4	5	6	7	8	9	10
27	6.585	5.119	4.293	3.753	3.368	3.077	2.847	2.661	2.507
28	6.608	5.141	4.315	3.774	3.388	3.096	2.867	2.680	2.525
29	6.630	5.162	4.335	3.794	3.407	3.115	2.885	2.698	2.542
30	6.650	5.182	4.354	3.813	3.426	3.133	2.902	2.715	2.559
40	6.801	5.331	4.500	3.955	3.565	3.270	3.037	2.847	2.689
50	6.894	5.424	4.592	4.047	3.657	3.361	3.127	2.937	2.778
60	6.957	5.487	4.656	4.111	3.721	3.425	3.192	3.001	2.842
80	7.038	5.569	4.739	4.195	3.806	3.511	3.277	3.087	2.929
100	7.087	5.620	4.791	4.247	3.859	3.564	3.332	3.143	2.985
120	7.120	5.654	4.826	4.283	3.895	3.602	3.370	3.181	3.024
160	7.161	5.697	4.870	4.329	3.942	3.649	3.419	3.231	3.074
200	7.187	5.723	4.897	4.357	3.971	3.679	3.449	3.262	3.106
300	7.220	5.759	4.934	4.395	4.010	3.719	3.490	3.304	3.150
∞*	7.289	5.831	5.009	4.473	4.091	3.803	3.577	3.394	3.242

* C = Entry value * $(k/2)$

ference. Conaway et al. (1991) have given such results for five tests, so we compute results only for the ALR test. As with the LR and circular cone tests, the computation method is simulation with 2,000 replications (with standard error less than .012); for the other tests, computations were done directly. In all cases σ^2 is known to be 1 and all $n_i = 1$. Two directions that tend to attain the maximum and the minimum power for all the tests are considered: the center direction $(-k, 1, \ldots, 1)$ and the edge direction $(-1, k, -1, \ldots, -1)$. Along each direction, the parameter vector is specified by giving its squared distance from the origin, $\Delta^2 = \sum(\theta_i - \bar{\theta})^2$, where $\bar{\theta} = (k+1)^{-1}\sum\theta_i$.

Table 2 indicates that the LR test is the minimax choice among the six tests, which means that the maximum type II error for the alternatives of a given distance from the null space is minimized or, equivalently, the minimum power is maximized. Among the other tests, the circular cone test is closest to, but seems to be dominated by, the LR test. The ALR test is second closest. Because its cone is contained in that of the LR test, it is more powerful in the center and less powerful in the edge. The gain in the center is much bigger than the loss in the edge in the high-dimensional cases $k = 5$ and 9, making the ALR test an attractive alternative. Dunnett's test is close to the ALR test, but its maximum power is higher and its minimum power is lower because its contrasts form a smaller cone. In this power trend the extreme is the single-contrast test, which has the maximum possible power in the center. With the minimum power in the center of all tests, the OC test is the extreme in the opposite trend. This disadvantage is balanced by the advantage that the power in the edge tends to be the maximum of all tests. In the case $k = 2$ the OC test even looks closer to

Table 2. Power Comparison Between ALR and Five Other Tests

k	Direction	Δ	Test					
			ALR	LR	Circular	Optimal	OC test	Dunnett
2	Center	1	23.05	21.80	21.80	25.95	20.10	23.77
		2	58.30	54.25	54.25	63.88	50.23	59.05
		3	87.00	85.65	85.65	91.23	81.19	88.13
	Edge	1	16.20	16.15	16.15	12.61	17.27	15.94
		2	46.85	48.65	48.65	25.95	49.44	43.07
		3	82.25	81.10	81.10	44.24	82.78	75.29
3	Center	1	20.90	19.45	19.10	25.95	17.29	23.07
		2	55.35	50.25	50.00	63.88	42.18	57.36
		3	86.35	81.60	81.15	91.23	72.36	86.87
	Edge	1	12.50	15.35	15.00	9.48	14.13	12.58
		2	39.65	41.05	39.95	16.40	42.91	34.13
		3	73.05	76.35	74.80	25.95	77.96	65.24
5	Center	1	20.20	15.70	14.90	25.95	14.41	22.55
		2	52.00	41.65	39.70	63.88	33.08	56.11
		3	83.45	74.60	73.10	91.23	59.23	85.88
	Edge	1	10.70	11.25	11.35	7.42	11.39	9.83
		2	29.40	31.35	29.90	10.66	36.22	25.84
		3	63.55	65.85	62.70	14.80	72.24	53.87
9	Center	1	19.25	14.95	14.00	25.95	11.90	22.38
		2	48.30	35.90	33.55	63.88	24.76	55.65
		3	80.60	66.40	64.35	91.23	44.25	85.49
	Edge	1	8.25	9.75	10.35	6.25	9.31	7.95
		2	23.05	26.35	23.90	7.74	30.01	19.38
		3	51.45	55.00	52.15	9.48	65.97	43.33

be the minimax choice than the LR test. In the cases $k = 5$ and 9, the power is higher in the edge than in the center. From a power standpoint we conclude that the LR, circular cone, and ALR tests are more appropriate for general use than the other tests.

4. CHOICE OF SAMPLE SIZE

Here we consider an appropriate choice of sample size to use with the ALR test. The first situation is that N is fixed and $n_1 = \cdots = n_k$ $(= n)$, so the focus is on the ratio n_0/n. No design is expected to make the ALR test uniformly more powerful than under any other design. Therefore, a subjective criterion must be used. We attempted several optimization approaches, but the results varied widely, so we also conducted a simulation study. The results, given in Table 3, are the basis for the following recommendations. A design whose ratio n_0/n is greater than k or smaller than 1 should be avoided (see the case $k = 2$ in the table). The ratio 1 seems close to being minimax, but the ratio $k^{1/2}$, which Dunnett (1955) recommended for his test, is preferable for a large gain in the center and only a small loss in the edge.

In the second design situation, the power at a particular point is required to be at least a given number γ. For simplicity, assume that $n_0 = \cdots = n_k = n$ and σ^2 is known to be 1, so n is to be determined. The particular point is chosen by specifying its (usual) distance from the null hypothesis, say δ, and its direction, say $\mathbf{e}_1$ in the $\boldsymbol{\mu}$ domain. This leads to $\boldsymbol{\mu}^* = n^{1/2}\delta^{1/2}\mathbf{e}_1/(\mathbf{e}_1'\boldsymbol{\Omega}^{-1}\mathbf{e}_1)^{1/2}$. Because σ^2 is known, the ALR test statistic is $\hat{\boldsymbol{\mu}}'\boldsymbol{\Omega}^{-1}\hat{\boldsymbol{\mu}} = \sum(w_i^+)^2$. At the given point, we have $\mathbf{w} \sim N(\nu^*, I)$, where $\nu^* = \mathbf{A}\boldsymbol{\mu}^*$. Observe that $\mathbf{A}\mathbf{e}_1 = n^{1/2}\{\mathbf{e}_1 - k^{-1}[1-(k+1)^{-1/2}]\mathbf{J}\}$ has only one positive component, the magnitude of which is much larger than the magnitudes of the negative components. This suggests that the contribution from w_i, $i > 1$ to the sum $\sum(w_i^+)^2$ may be negligible. Therefore, instead of controlling $P\{\sum(w_i^+)^2 > c\}$, we control $P\{(w_1^+)^2 > c\} = P\{w_1 > c^{1/2}\}$ to be γ, which is easily done because w_1 is a normal variable with mean $n^{1/2}\delta^{1/2}\{(k-1)(k+1)^{1/2}+1\}k^{-3/2}$ and variance 1. Note that the sample size thus obtained is an upper bound to the required size. We use an example of Mukerjee et al. (1987) to demonstrate the accuracy of the foregoing approximation. Here $k = 5$ and $\boldsymbol{\theta} = (30)^{1/2}(-1, 5, -1, -1, -1, -1)$. The test level is .05, and the power at this $\boldsymbol{\theta}$ is .8. By Table 1, $c = 7.48$. The mean of w_1 is calculated to be $.9658(n)^{1/2}$. Using a normal table, we obtain $n = 13.7$; the final solution is $n = 14$. Mukerjee et al. (1987) gave $n = 16$ for Dunnett's test, $n = 11$ for the OC test, and $n = 13$ for the LR test. In view of the relative power performance shown in Table 2 the estimate $n = 14$ for the ALR test seems quite close to that required. By simulation, the power for $n = 14$ is about .83, and the power for $n = 13$ is about .8; the actual value for the latter case may be lower than .8. Simulation under unknown σ^2 leads to .83 for $n = 14$. Note that although the OC test has the smallest n, the power achieved at the given point is not the minimum power over all directions.

5. CLOSED TESTING

In the multivariate framework, any test, however powerful, can only conclude that there exist one or more new treatments that are better than the control. To identify the better treatments, the closed testing procedure of Marcus et al. (1976) can be used. To this end, we define a null hypothesis for each nonempty subset of the new treatments, which states that the treatments in this subset are all equivalent to the control. The closed testing procedure enables all of these null hypotheses to be tested in a framework that controls the overall type I error, say to be α. The procedure is dictated by two rules. First, if a hypothesis is to be actually tested, then the level is α. The test method can vary from one hypothesis to another, but in our discussion it is the same for all hypotheses. Second, a hypothesis is actually tested only if every other hypothesis that implies it has actually been tested and rejected. This means that any hypothesis not actually tested is automatically accepted. Operationally, one must start with the global hypothesis; that is, the original null hypothesis. The procedure continues only if this is rejected. Then a new treatment is deleted and the test method is applied again, and so on. Although all individual test levels are α, the overall level remains α. If a treatment is sufficiently better than the control, then all of those hypotheses that contain this treatment will be rejected (by a reasonable method), leading to the identification of the treatment. It is possible that the global hypothesis is rejected but all pairwise comparison hypotheses are accepted.

Table 4 shows simulation results comparing the ALR, LR, and Dunnett tests in closed testing, focusing on their ability to identify the better treatments. Note that although Dunnett's test can directly make pairwise comparisons, its use in closed testing as a multivariate test results in a more powerful procedure (Marcus et al. 1976). For simplicity, we

Table 3. Power of λ_{ALR} Under Different Designs

k	Direction	Δ	Design (n_0, n)				
			(14, 8)	(12, 9)	(10, 10)	(8, 11)	(6, 12)
2	Center	.3	51.1	50.1	47.5	42.4	37.9
		.5	69.1	67.6	64.7	60.5	53.5
		.8	86.6	84.1	81.1	76.6	69.6
	Edge	.3	35.2	35.9	36.4	36.3	35.5
		.5	50.9	52.4	52.8	52.7	52.1
		.8	69.3	70.9	71.9	72.2	71.3
			(22, 6)	(19, 7)	(16, 8)	(13, 9)	(10, 10)
3	Center	.1	25.5	25.6	25.1	23.7	22.3
		.5	70.7	71.2	70.0	67.0	61.6
		.8	87.5	87.8	86.5	83.6	77.5
	Edge	.1	12.6	12.9	13.1	13.2	13.5
		.5	35.4	38.8	39.9	41.4	41.9
		.8	50.4	55.2	58.7	61.3	62.3
			(39, 9)	(31, 11)	(27, 12)	(23, 13)	(15, 15)
4	Center	.1	37.3	36.8	35.7	33.5	28.6
		.3	72.3	71.4	69.2	65.5	55.2
		.5	89.5	88.9	87.5	85.1	75.3
		.8	98.1	97.9	97.5	96.4	90.4
	Edge	.1	13.9	15.0	15.2	15.5	15.5
		.3	32.2	35.7	36.7	38.2	38.8
		.5	46.5	50.6	52.6	54.4	56.8
		.8	64.5	70.9	74.1	75.8	78.9

Table 4. Closed Testing Power Comparison Among ALR, Dunnett, and LR Tests, $k = 5$

Direction	Δ^2	Test	Hypothesis: Global	Pairwise				
Center	8	ALR	77.9	53.1	53.6	52.6	50.9	53.7
		Dunnett	82.0	55.3	55.7	54.4	52.7	56.5
		LR	67.6	45.4	46.5	46.1	44.2	46.6
	9	ALR	82.2	59.9	58.8	60.0	57.3	59.6
		Dunnett	85.5	60.8	60.4	61.6	58.9	62.1
		LR	73.6	52.2	53.1	53.1	50.2	53.0
Edge	13	ALR	80.2	75.6	1.2	1.4	1.3	1.3
		Dunnett	70.3	70.3	1.4	1.5	1.7	1.5
		LR	83.4	73.6	1.1	.8	1.0	1.0
	14	ALR	84.1	79.4	1.2	1.4	1.3	1.3
		Dunnett	73.3	73.3	1.4	1.5	1.7	1.5
		LR	87.1	76.8	1.1	.8	1.0	1.0

report only rejection rates of the global hypothesis and of those of the pairwise comparisons, based on 1,000 replications (with standard error below .016). Comparing the ALR test to Dunnett's test, the difference at the pairwise comparison level is in the same direction as, but is smaller than, that at the global level. Comparing the ALR test to the LR test, the trend unexpectedly reverses in the edge direction. However, in other cases that are not shown here and where the global power is low, say below .6, the reverse does not occur.

6. CONCLUSION

Although the ALR test is developed as an approximation to the LR test, it is an exact test with a theory that requires no constraint on the sample sizes. It is simple in that (a) the critical value is in most cases available from Table 1 and, if necessary, can easily be computed; (b) computing the statistic is elementary in the case of equal treatment sample sizes; and (c) the total sample size can be calculated based on a normal probability table. For the LR test, one needs a computer program to compute the critical value even when all treatment sample sizes are the same but are not equal to that of the control. In the case of unequal treatment sample sizes, computing the ALR statistic requires a computer for some matrix operations. But this task is easier than computing multivariate normal integrals required by the LR test. In terms of power the ALR test is also appealing, in that its minimum power seems only slightly lower but its maximum power can be substantially higher. It also has a power advantage in the closed testing setting. In conclusion, the ALR test provides a practically feasible way to implement the LR test. Compared to the circular cone test, the ALR test also has a clear edge in simplicity with only a small loss in minimum power. In practice the choice of test depends much on one's perception of the importance of simplicity relative to power, but the richness of available tests makes it easy to find a test that matches the particular perception.

APPENDIX: A PROCEDURE FOR OBTAINING THE **A** MATRIX

Here is the procedure of Tang et al. (1989) for obtaining a matrix $\mathbf{A}$ satisfying

$$\text{(a)}\ \mathbf{A}'\mathbf{A} = \boldsymbol{\Omega}^{-1}$$

and

$$\text{(b)}\ \mathbf{J}'\mathbf{A}\mathbf{D} = p\mathbf{J}' \quad \text{for} \quad \text{a scalar } p > 0,$$

where $\boldsymbol{\Omega}$ is any given positive definite matrix and $\mathbf{D}$ is a diagonal matrix whose diagonal elements are $(\mathbf{e}_i'\boldsymbol{\Omega}^{-1}\mathbf{e}_i)^{-1/2}$.

Step 1. Compute the Choleski decomposition, $\boldsymbol{\Omega}^{-1} = \mathbf{C}'\mathbf{C}$, with $\mathbf{C}$ upper triangular. Then compute $\mathbf{d} = \mathbf{C}'^{-1}\mathbf{D}^{-1}\mathbf{J}$.

Step 2. Compute two orthogonal matrices, $\mathbf{Q}_1$ and $\mathbf{Q}_2$, by applying the Gram–Schmidt orthonormalization process to the vectors $\mathbf{d}, \mathbf{e}_2, \ldots, \mathbf{e}_k$, and to the vectors $\mathbf{J}, \mathbf{e}_2, \ldots, \mathbf{e}_k$. Then compute $\mathbf{A} = \mathbf{Q}_2\mathbf{Q}_1'\mathbf{C}$.

[Received December 1994. Revised August 1996.]

REFERENCES

Abelson, R. P., and Tukey, J. W. (1963), "Efficient Utilization of Non-Numerical Information in Quantitative Analysis: General Theory and the Case of the Simple Order," *Annals of Mathematical Statistics*, 34, 1347–1369.

Akkerboom, J. C. (1990), *Testing Problems With Linear or Angular Inequality Constraints*, New York: Springer-Verlag.

Aptech Systems, Inc. (1991), *GAUSS User's Manual*, Version 2.1, Kent, WA: Author.

Bartholomew, D. J. (1961), "A Test for Homogeneity of Means Under Restricted Alternatives," *Journal of the Royal Statistical Society*, Ser. B, 23, 239–281.

Conaway, M., Pillers, C., Robertson, T., and Sconing, J. (1991), "A Circular-Cone Test for Testing Homogeneity Against a Simple Tree Order," *The Canadian Journal of Statistics*, 19, 283–296.

Dunnett, C. W. (1955), "A Multiple Comparisons Procedure for Comparing Several Treatments With a Control," *Journal of the American Statistical Association*, 50, 1096–1121.

Dunnett, C. W., and Tamhane, A. C. (1991), "Step-Down Multiple Tests for Comparing Treatments With a Control in Unbalanced One-Way Layouts," *Statistics in Medicine*, 10, 939–947.

——— (1992), "A Step-Up Multiple Test Procedure," *Journal of the American Statistical Association*, 87, 162–170.

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.

Mukerjee, H., Robertson, T., and Wright, F. T. (1987), "Comparison of Several Treatments With a Control Using Multiple Contrasts," *Journal of the American Statistical Association*, 82, 902–910.

Robertson, T., Wright, F. T., and Dykstra, R. L. (1988), *Order-Restricted Statistical Inference*, New York: Wiley.

Schaafsma, W., and Smid, L. J. (1966), "Most Stringent Somewhere Most Powerful Tests Against Alternatives Restricted by a Number of Linear Inequalities," *Annals of Mathematical Statistics*, 37, 1161–1172.

Steel, R. G. D. (1959), "A Multiple Comparison Rank Sum Test: Treatments Versus Control," *Biometrics*, 15, 560–572.

Tang, D.-I., Geller, N. L., and Pocock, S. J. (1993), "On the Design and Analysis of Randomized Clinical Trials With Multiple Endpoints," *Biometrics*, 49, 23–30.

Tang, D.-I., Gnecco, C., and Geller, N. L. (1989), "An Approximate Likelihood Ratio Test for a Normal Mean Vector With Nonnegative Components With Application to Clinical Trials," *Biometrika*, 76, 577–583.

Tang, D.-I., and Lin, S. P. (1994), "On Improving Some Methods for Multiple Endpoints," in *Proceedings of the Biopharmaceutical Section, American Statistical Association.*