Accurate Critical Constants for the One-Sided Approximate Likelihood Ratio Test of a Normal Mean Vector When the Covariance Matrix Is Estimated

Ajit C. Tamhane   and   Brent R. Logan

Department of Statistics   Division of Biostatistics

Northwestern University   Medical College of Wisconsin

2006 Sheridan Road   8701 Watertown Plank Rd.

Evanston, IL 60208, USA   Milwaukee, WI 53226, USA

E-mail: ajit@iems.northwestern.edu   E-mail: blogan@mcw.edu

SUMMARY. Tang, Gnecco and Geller (1989) proposed an approximate likelihood ratio (ALR) test of the null hypothesis that a normal mean vector equals a null vector against the alternative that all of its components are nonnegative with at least one strictly positive. This test is useful for comparing a treatment group with a control group on multiple endpoints, and the data from the two groups are assumed to follow multivariate normal distributions with different mean vectors and a common covariance matrix (the homoscedastic case). Tang et al. derived the test statistic and its null distribution assuming a known covariance matrix. In practice, when the covariance matrix is estimated, the critical constants tabulated by Tang et al. result in a highly liberal test. To deal with this problem, we derive an accurate small sample approximation to the null distribution of the ALR test statistic by using the moment matching method. The proposed approximation is then extended to the heteroscedastic case. The accuracy of both the approximations is verified by simulations. A real data example is given to illustrate the use of the approximations.

KEY WORDS: Chi-bar squared distribution; Heteroscedastic; Homoscedastic; Hotelling's T^2 distribution; Multiple endpoints; Multivariate normal distribution.

1. Introduction

Many clinical trials are conducted to compare a treatment group with a control group on several endpoints. Often, the treatment is expected to have a positive effect on all endpoints. Inappropriateness of Hotelling's T^2 test for this problem has been noted by several authors; see, e.g., Meier (1975). As an alternative to Hotelling's T^2 test, O'Brien (1984) proposed his OLS and GLS tests, which possess high power against the restricted alternative that the treatment has the same positive standardized effect on all endpoints. In the one-sample problem, for the more general alternative that the mean vector lies in the positive orthant, the exact likelihood ratio (LR) tests were derived by Kudô (1963) assuming a known covariance matrix and by Perlman (1969) assuming an unknown covariance matrix. However, these test statistics are complicated, and their null distributions are difficult to obtain. For the Perlman test statistic, the null distribution is not free of the unknown covariance matrix. To obviate some of these difficulties, Tang, Gnecco, and Geller (1989) proposed an approximate likelihood ratio (ALR) test for the one sample problem assuming a known covariance matrix. It provides an easy to use approximation to Kudô's test. When extended to the two-sample problem, it has better power properties compared to O'Brien's (1984) OLS and GLS tests for most alternatives in the positive orthant. This note focuses on the ALR test.

In practical applications the population covariance matrix is always unknown, and the sample covariance matrix must be used to estimate it. This results in a highly liberal ALR test if one uses the null distribution derived by Tang et al. which assumes a known covariance matrix; see Reitmeir and Wassmer (1996) and Sankoh et al. (1999). The liberalism decreases as the degrees of freedom (d.f.) available to estimate the covariance matrix increases. For example, for six endpoints, the estimated type I error rate for a nominal 0.05-level test is 0.3550 for 10 d.f., 0.1066 for 30 d.f., 0.0830 for 50 d.f. and 0.0611 for 100 d.f. We derive an accurate approximation to the small sample distribution of the ALR test which eliminates

this liberalism.

The outline of the paper is as follows. In Section 2 we set the notation and review the ALR test for the two sample problem in the known common covariance matrix (homoscedastic) case. In Section 3 we propose the approximation to its small sample null distribution in the unknown covariance matrix case. In Section 4 we extend the approximation to the unequal unknown covariance matrices (heteroscedastic) case. In Section 5 we present the results of simulations that demonstrate the accuracy of the proposed approximations. An example is given in Section 6. Finally, a discussion about some anomalies of the LR tests (including the ALR test) is given in Section 7.

2. Notation, Problem Formulation and the ALR Test

Suppose that there are two independent treatment groups with n_1 and n_2 subjects on each of whom $m \geq 2$ endpoints are measured. Treatment 1 is the test treatment and treatment 2 is the control. Let x_{ijk} denote the measurement on the kth endpoint for the jth subject in the ith treatment group. For treatment group i, assume that $\boldsymbol{x}_{ij} = (x_{ij1}, x_{ij2}, \ldots, x_{ijm})'$, $j = 1, 2, \ldots, n_i$, are independent and identically distributed (i.i.d.) random vectors from an m-variate normal distribution with mean vector $\boldsymbol{\mu}_i = (\mu_{i1}, \mu_{i2}, \ldots, \mu_{im})'$ and covariance matrix $\boldsymbol{\Sigma}_i$ ($i = 1, 2$). In the present and the next section we assume the homoscedastic case, i.e., $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}$ (say). In this section we assume that $\boldsymbol{\Sigma}$ is known.

Let $\boldsymbol{\delta} = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = (\delta_1, \delta_2, \ldots, \delta_m)'$ denote the vector of mean differences. We are interested in testing the null hypothesis of no difference against the one-sided alternative:

$$H_0 : \boldsymbol{\delta} = \mathbf{0} \text{ vs. } H_1 : \boldsymbol{\delta} \in \mathcal{O}^+, \tag{2.1}$$

where $\mathbf{0}$ is the null vector and $\mathcal{O}^+ = \{\boldsymbol{\delta}|\delta_k \geq 0 \text{ for } k = 1, 2, \ldots, m, \boldsymbol{\delta} \neq \mathbf{0}\}$ is the positive orthant.

Let $\overline{\boldsymbol{x}}_{i\cdot} = (\overline{x}_{i\cdot1}, \overline{x}_{i\cdot2}, \ldots, \overline{x}_{i\cdot m})'$ denote the vector of sample means of the n_i subjects from

the ith group ($i = 1, 2$). The first step in the ALR test is to compute the transformation

$$\boldsymbol{u} = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \boldsymbol{A} (\overline{\boldsymbol{x}}_{1.} - \overline{\boldsymbol{x}}_{2.}), \tag{2.2}$$

where $\boldsymbol{A}$ is any positive definite matrix such that

$$\boldsymbol{A}' \boldsymbol{A} = \boldsymbol{\Sigma}^{-1} \text{ and } \boldsymbol{A} \boldsymbol{\Sigma} \boldsymbol{A}' = \boldsymbol{I}. \tag{2.3}$$

Then $\boldsymbol{u} \sim N(\boldsymbol{\theta}, \boldsymbol{I})$, where $\boldsymbol{\theta} = \{n_1 n_2 / (n_1 + n_2)\}^{1/2} \boldsymbol{A} \boldsymbol{\delta}$ and the hypotheses (2.1) become $H_0 : \boldsymbol{\theta} = \boldsymbol{0}$ vs. $H_1 : \boldsymbol{\theta} \in \boldsymbol{A}(\mathcal{O}^+)$, where $\boldsymbol{A}(\mathcal{O}^+) = \left[\{n_1 n_2 / (n_1 + n_2)\}^{1/2} \boldsymbol{A} \boldsymbol{\delta} \middle| \boldsymbol{\delta} \in \mathcal{O}^+ \right]$ is a polyhedral cone. The matrix $\boldsymbol{A}$ used in the transformation is not unique. Tang et al. gave a method for choosing $\boldsymbol{A}$ such that the center direction of $\boldsymbol{A}(\mathcal{O}^+)$ coincides with the center direction of $\mathcal{O}^+$ which is the equicoordinate ray $(\lambda, \lambda, \ldots, \lambda)'$ for $\lambda > 0$.

The cone alternative $\boldsymbol{A}(\mathcal{O}^+)$ is approximated by $\mathcal{O}^+$. Then the ALR test statistic equals

$$g(\boldsymbol{u}) = \sum_{k=1}^{m} \{\max(u_k, 0)\}^2. \tag{2.4}$$

The null distribution of $g(\boldsymbol{u})$ is the $\overline{\chi}^2$ distribution (see Robertson, Wright, and Dykstra, 1988) with symmetric binomial probability weights:

$$\Pr\{g(\boldsymbol{u}) > c\} = \sum_{k=0}^{m} \binom{m}{k} 2^{-m} \Pr\left(\chi_k^2 > c\right), \tag{2.5}$$

where $\chi_0^2 = 0$.

3. Small Sample Null Distribution of the ALR Test Statistic in the Homoscedastic Case

In this section we assume that $\boldsymbol{\Sigma}$ is unknown. Let $\widehat{\boldsymbol{\Sigma}}$ denote the pooled covariance matrix with $\nu = n_1 + n_2 - 2$ d.f. It is known that $\widehat{\boldsymbol{\Sigma}}$ is positive definite with probability 1 (Eaton and Perlman, 1973). Analogous to the definition of $\boldsymbol{u}$, we first make the transformation

$$\boldsymbol{v} = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \boldsymbol{B} (\overline{\boldsymbol{x}}_{1.} - \overline{\boldsymbol{x}}_{2.}), \tag{3.1}$$

where $\boldsymbol{v} = (v_1, v_2, \ldots, v_m)'$ and $\boldsymbol{B}$ is a positive definite matrix such that

$$\boldsymbol{B}'\boldsymbol{B} = \widehat{\boldsymbol{\Sigma}}^{-1} \text{ and } \boldsymbol{B}\widehat{\boldsymbol{\Sigma}}\boldsymbol{B}' = \boldsymbol{I}. \tag{3.2}$$

Analogous to (2.4), the ALR test statistic is given by

$$g(\boldsymbol{v}) = \sum_{k=1}^{m} \left\{ \max(v_k, 0) \right\}^2. \tag{3.3}$$

The choice of $\boldsymbol{B}$ is not unique and, depending on the choice, the null distribution of $\boldsymbol{v}$ may or may not depend on $\boldsymbol{\Sigma}$. In particular, it is straightforward to show that if $\boldsymbol{B}$ is the unique lower triangular matrix obtained using the Cholesky decomposition then the null distribution of $\boldsymbol{v}$ does not depend on $\boldsymbol{\Sigma}$. Also, it is easy to see that this distribution is symmetric about the origin, and as $\nu \to \infty$, it approaches the joint distribution of $\boldsymbol{u}$, the components of which are i.i.d. $N(0,1)$ random variables (r.v.'s).

The null distribution of $g(\boldsymbol{v})$ is quite intractable. Therefore we derive an approximation to it. For finite ν, by analogy to the $\overline{\chi}^2$ distribution (2.5), we propose to approximate

$$\Pr\{g(\boldsymbol{v}) > c\} \approx \sum_{k=0}^{m} \binom{m}{k} 2^{-m} \Pr\left\{ \left(\frac{\nu k}{\nu - m + 1} \right) F_{k, \nu - m + 1} > c \right\}, \tag{3.4}$$

where $F_{0, \nu - m + 1} = 0$. Thus we approximate the distribution of $g(\boldsymbol{v})$ by a mixture of the distributions of

$$\left(\frac{\nu k}{\nu - m + 1} \right) F_{k, \nu - m + 1}, \quad k = 0, 1, \ldots, m,$$

with symmetric binomial probability weights. We will call this approximation as the $\overline{F}$ approximation since it uses a mixture of scaled F r.v.'s. This approximation is exact for all ν when $m = 1$ and for $\nu = \infty$ when $m > 1$.

We now show that the first moments of $g(\boldsymbol{v})$ and the $\overline{F}$ approximation match. To compute $E\{g(\boldsymbol{v})\}$ we use the fact that

$$\sum_{k=1}^{m} v_k^2 = \boldsymbol{v}'\boldsymbol{v} \sim T_{m,\nu}^2 \sim \left(\frac{\nu m}{\nu - m + 1} \right) F_{m, \nu - m + 1}, \tag{3.5}$$

where $T^2_{m,\nu}$ is Hotelling's T^2 random variable (Anderson, 1984). Since each v_k is symmetrically distributed around 0, we have (assuming that $\nu > m + 1$)

$$E\{g(\boldsymbol{v})\} = \sum_{k=1}^{m} E\{\max(v_k, 0)\}^2 = \sum_{k=1}^{m} \frac{1}{2} E(v_k^2) = \frac{1}{2} E(\boldsymbol{v}'\boldsymbol{v}) = \frac{1}{2} E(T^2_{m,\nu}) = \frac{\nu m}{2(\nu - m - 1)}.$$

Next, to compute the expected value of the $\overline{F}$ approximation, define a r.v. $x = \{\nu y/(\nu - m + 1)\} F_{y,\nu-m+1}$, where y is a binomial r.v. with sample size m and success probability $1/2$. Then $E(x) = E\{E(x|y)\}$ equals

$$E\left\{\frac{\nu y}{\nu - m + 1} E(F_{y,\nu-m+1})\right\} = E\left(\frac{\nu y}{\nu - m + 1} \times \frac{\nu - m + 1}{\nu - m - 1}\right) = \frac{\nu E(y)}{\nu - m - 1} = \frac{\nu m}{2(\nu - m - 1)}.$$

Therefore the approximation matches with the exact distribution on the first moment.

We also compared the second moments of $g(\boldsymbol{v})$ and the $\overline{F}$ approximation. It can be shown that the exact variance of the $\overline{F}$ distribution is given by

$$m\left(\frac{\nu}{\nu - m - 1}\right)^2 \left\{\frac{1}{4} + \frac{2\nu - m - 1}{2(\nu - m - 3)}\right\}$$

and an approximation to the variance of $g(\boldsymbol{v})$ is given by

$$m\left(\frac{\nu}{\nu - m - 1}\right)^2 \left[\left(\frac{\nu - m - 1}{\nu - 2}\right)^2 \left\{\frac{1}{4} + \frac{\nu - 1}{2(\nu - 4)}\right\} + \frac{\nu - 1}{2(\nu - m - 3)}\right].$$

The details of the derivations of these expressions are given in Logan (2001). For $m = 1$, the two expressions are equal to

$$\left(\frac{\nu}{\nu - 2}\right)^2 \left(\frac{1}{4} + \frac{\nu - 1}{\nu - 4}\right),$$

which is the variance of $\{\max(v_1, 0)\}^2$ since $v_1 \sim t_\nu$ for $m = 1$. (Note that the v_k are not marginally t-distributed for $m > 1$.) Also, as $\nu \to \infty$, both expressions approach $5m/4$, which is the variance of the $\overline{\chi}^2$ distribution. For other cases, the variance of the $\overline{F}$ distribution is found to be higher, suggesting that the $\overline{F}$ approximation will be conservative. Critical constants for the $\overline{F}$ approximation are given in Table 1 for $\alpha = 0.01, 0.05$ and 0.10.

4. Small Sample Null Distribution of the ALR Test Statistic in the Heteroscedastic Case

In what follows, it will be convenient to define $\boldsymbol{\Omega}_i = (1/n_i)\boldsymbol{\Sigma}_i \ \ (i = 1, 2), \boldsymbol{\Omega} = \boldsymbol{\Omega}_1 + \boldsymbol{\Omega}_2$ and $\boldsymbol{\Sigma} = \{n_1 n_2/(n_1 + n_2)\}\boldsymbol{\Omega}$. The sample estimates of these matrices are denoted by putting carets over them; thus $\widehat{\boldsymbol{\Sigma}}_i$ denotes the sample covariance matrix from group i with $n_i - 1$ d.f.

Analogous to (2.2) and (3.1), we now have the transformation

$$\boldsymbol{w} = \sqrt{\frac{n_1 n_2}{n_1 + n_2}}\boldsymbol{C}(\overline{\boldsymbol{x}}_{1\cdot} - \overline{\boldsymbol{x}}_{2\cdot}), \tag{4.1}$$

where $\boldsymbol{w} = (w_1, w_2, \ldots, w_m)'$ and $\boldsymbol{C}$ is a positive definite matrix such that $\boldsymbol{C}'\boldsymbol{C} = \widehat{\boldsymbol{\Sigma}}^{-1}$ and $\boldsymbol{C}\widehat{\boldsymbol{\Sigma}}\boldsymbol{C}' = \boldsymbol{I}$. Analogous to (3.3), the ALR test statistic is given by

$$g(\boldsymbol{w}) = \sum_{k=1}^{m}\left\{\max(w_k, 0)\right\}^2. \tag{4.2}$$

We propose the same $\overline{F}$ distribution (3.4) as an approximation to the null distribution of $g(\boldsymbol{w})$, but with the following Welch-Satterthwaite estimate of the d.f. ν derived by Yao (1965) for the multivariate Behrens-Fisher problem:

$$\frac{1}{\nu} = \frac{1}{(\boldsymbol{d}'\widehat{\boldsymbol{\Omega}}^{-1}\boldsymbol{d})^2}\left[\frac{(\boldsymbol{d}'\widehat{\boldsymbol{\Omega}}^{-1}\widehat{\boldsymbol{\Omega}}_1\widehat{\boldsymbol{\Omega}}^{-1}\boldsymbol{d})^2}{n_1 - 1} + \frac{(\boldsymbol{d}'\widehat{\boldsymbol{\Omega}}^{-1}\widehat{\boldsymbol{\Omega}}_2\widehat{\boldsymbol{\Omega}}^{-1}\boldsymbol{d})^2}{n_2 - 1}\right], \tag{4.3}$$

where $\boldsymbol{d} = (\overline{\boldsymbol{x}}_{1\cdot} - \overline{\boldsymbol{x}}_{2\cdot})$. Note that Yao derived this formula (also using the moment matching method) to approximate the distribution of $\boldsymbol{w}'\boldsymbol{w} = \{n_1 n_2/(n_1 + n_2)\}(\overline{\boldsymbol{x}}_{1\cdot} - \overline{\boldsymbol{x}}_{2\cdot})'\widehat{\boldsymbol{\Sigma}}^{-1}(\overline{\boldsymbol{x}}_{1\cdot} - \overline{\boldsymbol{x}}_{2\cdot})$ by Hotelling's $T_{m,\nu}^2 = \{\nu m/(\nu - m + 1)\}F_{m,\nu-m+1}$ distribution with an estimated ν, in analogy with the corresponding exact distribution result for $\boldsymbol{v}'\boldsymbol{v}$ with $\nu = n_1 + n_2 - 2$ given in (3.5). We have simply extended Yao's approximation to the $\overline{F}$ distribution.

5. Simulations

5.1 Homoscedastic Case

Simulations were designed to assess the actual type I error probabilities of the $\overline{F}$ approximation to the ALR test statistic distribution and compare it with the $\overline{\chi}^2$ approximation. Error probabilities were estimated for values of $m = 2(2)8$, $\nu = 10$, 30, and 50, and $\mathbf{\Sigma} = \mathbf{I}$. Additional simulations were done to study the dependence of $g(\boldsymbol{v})$ on $\mathbf{\Sigma}$ for two choices of $\mathbf{\Sigma}$ for $m = 4$ and $m = 8$. The first choice was an equicorrelated matrix with common correlation $\rho = 0.0, 0.3, 0.5, 0.7$ and 0.9. The second choice was a block correlated matrix with two blocks of size $m/2$; the within block and between block correlations (ρ_1 and ρ_2, respectively) were chosen to be $(\rho_1, \rho_2) = (0.5, 0.0), (0.9, 0.0)$ and $(0.9, 0.5)$. In both cases, two combinations of variances were examined: all $\sigma_i^2 = 1$, and half the $\sigma_i^2 = 1$ and the other half equal to 4.

In each of the 10,000 simulation runs, $\nu\widehat{\mathbf{\Sigma}}$ was sampled from a $W_m(\nu, \mathbf{\Sigma})$ random matrix, $\boldsymbol{u}$ was independently sampled with i.i.d. $N(0, 1)$ components, $\boldsymbol{v}$ was calculated using the relation $\boldsymbol{v} = \boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{u}$, and the ALR test statistic was computed using equation (3.3). The Tang et al. (1989) method was used to choose $\boldsymbol{A}$ and $\boldsymbol{B}$. The P-value was computed using the $\overline{\chi}^2$ and $\overline{F}$ approximations. The proportion of runs in which this P-value is $\leq \alpha$ gives an estimate of the type I error probability. The results for the $\mathbf{\Sigma} = \mathbf{I}$ case are given in Table 2. The results for the $\mathbf{\Sigma} \neq \mathbf{I}$ case were similar and hence are not reported to save space. In particular, the type I error probability of $g(\boldsymbol{v})$ is always well-controlled at or slightly below $\alpha = 0.05$. In no case did it exceed the upper 95% rejection limit of $0.05 + 1.96\{(0.05)(0.95)/10,000\}^{1/2} = 0.0543$. From these simulation results it appears that the null distribution of $g(\boldsymbol{v})$ does not depend in a significant way on $\mathbf{\Sigma}$. This observation

is bolstered by the fact that the first two moments of $g(\boldsymbol{v})$ (the second moment expression being approximate) given in Section 3 also do not depend on $\boldsymbol{\Sigma}$. Finally, the results show that the $\overline{\chi}^2$ approximation is liberal in all cases. For fixed ν, its liberalism increases as m increases.

5.2 Heteroscedastic Case

The type I error probability of the ALR test was simulated for nominal $\alpha = .05$, $m = 4$ and 8, and $n_1 = n_2 = 20, 30$ and 50. For each choice of m and $n_1 = n_2$, a total of eight combinations of $(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2)$ matrices were examined. These were parameterized as follows: Treatment Group: $(\boldsymbol{\Sigma}_1)_{ii} = \sigma_1^2$ $(1 \leq i \leq m/2)$, $(\boldsymbol{\Sigma}_1)_{ii} = \sigma_1'^2$ $(m/2 < i \leq m)$ and $(\boldsymbol{\Sigma}_1)_{ij} = \rho_1 (\boldsymbol{\Sigma}_1)_{ii}^{1/2} (\boldsymbol{\Sigma}_1)_{jj}^{1/2}$ $(1 \leq i \neq j \leq m)$; Control Group: $(\boldsymbol{\Sigma}_2)_{ii} = \sigma_2^2 = 1$ $(1 \leq i \leq m)$, $(\boldsymbol{\Sigma}_2)_{ij} = \rho_2$ $(1 \leq i \neq j \leq m)$.

All simulations were based on 10,000 runs. In each run, the P-value of the simulated $g(\boldsymbol{w})$ statistic was computed using the $\overline{F}$ approximation. The proportion of runs for which the P-value is $\leq \alpha$ gives an estimate of the type I error probability. The results are summarized in Table 3.

From these results we see that the type I error probability is well-controlled for $n_1 = n_2 = 30$ and 50, but for $n_1 = n_2 = 20$ there are several cases (especially for $m = 8$) where the estimated type I error probability exceeds the 0.05 level significantly; these cases are marked with asterisks. However, in no case does the estimated type I error probability exceed substantially, say, 0.06. Thus we conclude that the approximation is quite accurate for $n_1 = n_2 \geq 30$ and is acceptably accurate for $n_1 = n_2 = 20$ if $m \leq 8$. For larger values of m, these sample sizes may not be adequately large.

6. Example

Läuter, Kropf, and Glimm (1998) used data from a trial conducted by Dr. Michael Syn-

owitz at the Clinic for Heart Surgery of the Berlin Medical Faculty "CHARITE" in 1995. In this trial it was of interest to determine if autotransfusion (donation of the patient's own blood) makes surgical patients more sensitive to postoperative risk of infection (pneumonia, wound infection, etc.). This conjecture was supported by the fact that erythrocytes could be activated in the extracorporal circuit. Thirty patients about to undergo bypass surgery were randomized into two groups of 15 patients each. The treatment group received autotransfusion by means of a Pfizer-Shirley system, while the control group did not receive autotransfusion. As a measure of the risk, interleukin-6 (which is a pro-inflammation interleukin) plasma concentration (in picogram/ml) levels were measured for each patient at the time of surgical cut (t_0) and six successive occasions ($t_0 + 1$ hour, $t_0 + 3$ hours, $t_0 + 6$ hours, $t_0 + 12$ hours, $t_0 + 24$ hours and $t_0 + 48$ hours). Note that here we have repeated measures data on the same endpoint. This makes the assumption of one-sided effect more plausible than if we had different endpoints.

The first two measurements were dropped because of some missing and some inconsistent data, and the natural logarithmic transformation was applied to the remaining five measurements. The resulting data are shown in Table 4. The mean vectors and the sample covariance matrices for the two groups are shown below:

$$\overline{\boldsymbol{x}}_{1\cdot} = (3.50, 4.11, 3.77, 3.41, 3.15), \overline{\boldsymbol{x}}_{2\cdot} = (3.33, 3.55, 3.46, 2.85, 2.68),$$

$$\widehat{\boldsymbol{\Sigma}}_1 = \begin{bmatrix} 0.58 & 0.32 & 0.38 & 0.26 & 0.17 \\ & 0.49 & 0.44 & 0.33 & 0.21 \\ & & 0.53 & 0.41 & 0.21 \\ & & & 0.45 & 0.34 \\ & & & & 0.58 \end{bmatrix} \text{ and } \widehat{\boldsymbol{\Sigma}}_2 = \begin{bmatrix} 0.38 & 0.08 & 0.08 & 0.05 & -0.12 \\ & 0.21 & 0.00 & -0.08 & -0.09 \\ & & 0.22 & 0.13 & 0.02 \\ & & & 0.25 & 0.10 \\ & & & & 0.28 \end{bmatrix}.$$

Box's (1949) test for homogeneity of covariance matrices yielded a nonsignificant result ($\chi^2 = 16.958$ with a P-value $= 0.321$). Therefore $\widehat{\boldsymbol{\Sigma}}_1$ and $\widehat{\boldsymbol{\Sigma}}_2$ were pooled with a total

d.f. $= 28$ and the ALR test for the homoscedastic case was applied. The transformation matrix using the Tang et al. method is

$$\boldsymbol{B} = \begin{bmatrix} 1.57 & -0.77 & -0.53 & 0.16 & 0.74 \\ 0.07 & 2.11 & -1.44 & 0.77 & -0.11 \\ 0.03 & -0.27 & 2.72 & -2.25 & 0.55 \\ -0.13 & -0.21 & 0.10 & 2.05 & -1.42 \\ -0.85 & 0.02 & 0.37 & 0.50 & 1.00 \end{bmatrix}.$$

The ALR test statistic equals $g(\boldsymbol{v}) = 14.60$. Using the $\overline{\chi}^2$ approximation the P-value is 0.0022, whereas using the $\overline{F}$ approximation the P-value is 0.0145. Note the liberalism of the $\overline{\chi}^2$ approximation. The $\overline{F}$-approximated P-value is in line with the P-value of 0.0199 obtained using the standardized sum statistic of Läuter (1996) and 0.0104 obtained using the OLS statistic of O'Brien (1984).

For comparison purposes, the heteroscedastic ALR test was also performed. Because $n_1 = n_2$, $\widehat{\boldsymbol{\Sigma}}$ is the same in the heteroscedastic case as in the homoscedastic case. Therefore $\boldsymbol{C} = \boldsymbol{B}$ and $g(\boldsymbol{w}) = g(\boldsymbol{v}) = 14.60$. The d.f. using the formula (4.3) are $\nu = 25$ and the P-value $= 0.0172$. Thus the result is similar to that in the homoscedastic case.

7. Discussion

Silvapulle (1997) has given an example in the bivariate case where the LR test rejects H_0 even if both the mean differences are negative. This anomaly hinges on a large positive correlation between the endpoints. Many authors have noted other anomalies of the LR tests such as the lack of unbiasedness and monotonicity, and existence of uniformly more powerful tests. On the other hand, Perlman and Wu (1999) have defended LR tests (see, however, Perlman and Wu, 2000), noting that the alternative tests (e.g., those proposed by Berger (1989), Tang (1994), and Wang and McDermott (1998)) that are less biased and more powerful also suffer from lack of monotonicity and nonintuitive rejection regions.

They argued that the anomalies of LR tests are the result of an incorrectly specified null hypothesis. The anomalies disappear when the null hypothesis is specified as the complement of the one-sided alternative hypothesis.

Cohen and Sackrowitz (1998) have suggested cone ordered monotone tests to ameliorate these difficulties. However, their rejection regions are not entirely satisfactory either since, e.g., in the bivariate case, their test can reject H_0 if a large negative difference on one endpoint is overcome by a larger positive difference on the other endpoint; see Perlman and Wu (2001).

The abovementioned anomalies translate to the ALR test as well. Generally, they tend to occur when the endpoints are highly positively correlated. In this case, the endpoints with very dissimilar negative mean differences, when projected onto the positive orthant, can generate a large value of the LR statistic that warrants rejection.

It is not our intention to get into the pros and cons of tests for multiparameters hypothesis testing problems which seem to afflict LR as well as other tests. Suffice it to say that the above anomalies should not present a serious problem in most practical situations. When the correlation is high, it is unusual for the mean differences to be very dissimilar and the LR statistic to be large as a result. If a negative effect on some endpoint(s) cannot be ruled out *a priori* then a one-sided test should not be used. When a one-sided test is applicable, the ALR test is a good choice that performs better than the more popular OLS test. We have offered an accurate method for applying the ALR test in small samples. The proposed approximations are very simple to implement by using a C program available for download at `http://users.iems.northwestern.edu/∼ajit`.

acknowledge personal communications with Professors Joe Eaton, Mike Perlman and Shelby Haberman regarding the dependence of the null distribution of v on Σ.

REFERENCES

Anderson, T. W. (1984). *An Introduction to Multivariate Analysis*. Second Edition. New York: Wiley.

Berger, R. L. (1989). Uniformly more powerful tests for hypotheses concerning linear inequalities and normal means. *Journal of the American Statistical Association* **84**, 192– 199.

Box, G. E. P. (1949). A general distribution theory for a class of likelihood criteria. *Biometrika* **36**, 362– 389.

Cohen, A. and Sackrowitz, H. B. (1998). Directional tests for one-sided alternatives in multivariate models. *Annals of Statistics* **26**, 2321 -2338.

Eaton, M. L. and Perlman, M. D. (1973). The nonsingularity of generalized sample covariance matrices. *Annals of Statistics* **1**, 710 -717.

Follman, D. (1995). Multivariate tests for multiple endpoints in clinical trials. *Statistics in Medicine* **14**, 1163 -1175.

Kudô, A. (1963). A multivariate analogue of the one-sided test. *Biometrika* **50**, 403-418.

Logan, B. R. (2001). *Contributions to Multiple Endpoints and Dose Finding*. Doctoral Dissertation, Department of Statistics, Northwestern University, Evanston, IL.

Meier, P. (1975). Statistics and medical experimentation. *Biometrics* **31**, 511 -529.

O'Brien, P. C. (1984). Procedures for comparing samples with multiple endpoints. *Biometrics* **40**, 1079-1087.

Perlman, M. D. (1969). One-sided testing problems in multivariate analysis. *Annals of Mathematical Statistics* **40**, 549-567.

Perlman, M. D. and Wu, L. (1999). The emperor's new tests. *Statistical Science* **14**, 355–381.

Perlman, M. D. and Wu, L. (2000). On the validity of the likelihood ratio and maximum likelihood methods. *Statistical Science*, submitted.

Perlman, M. D. and Wu, L. (2001). A defense of the likelihood ratio criterion for testing one-sided and order-restricted alternatives. *Journal of Statistical Planning and Inference*, submitted.

Reitmeir, P. and Wassmer, G. (1996). One-sided multiple endpoint testing in two-sample comparisons. *Communications in Statistics (Computation and Simulation)* **25**, 99-117.

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

Sankoh, A. J., Huque, M. F., Russell, H. K. and D'Agostino, R. B., Sr. (1999). Global two-group multiple endpoint adjustment methods applied to clinical trials. *Drug Information Journal* **33**, 119–140.

Silvapulle, M. J. (1997). A curious example involving the likelihood ratio test against one-sided hypotheses. *The American Statistician* **51**, 178 -180.

Tang, D. I. (1994). Uniformly more powerful tests in a one-sided multivariate problem. *Journal of the American Statistical Association* **93**, 380 -386.

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.

Wang, Y. and McDermott, M. P. (1998). Conditional likelihood ratio test for a nonnegative normal mean vector. *Journal of the American Statistical Association* **89**, 380 -386.

Yao, Y. (1965). An approximate degrees of freedom solution to the multivariate Behrens-Fisher problem. *Biometrika* **52**, 139 -147.

Table 1: Critical Constants for ALR Test Based on the $\overline{F}$ Approximation

| | | m | | | | | | |
α	ν	2	3	4	5	6	7	8
0.10	10	4.00	6.49	9.80	14.63	22.47	37.14	72.05
	30	3.24	4.61	5.97	7.37	8.85	10.42	12.13
	50	3.12	4.35	5.52	6.67	7.83	9.01	10.23
	$\infty(\overline{\chi}^2)$	2.95	4.01	4.96	5.84	6.67	7.48	8.26
0.05	10	6.18	9.65	14.40	21.61	33.87	58.48	123.6
	30	4.75	6.41	8.05	9.74	11.52	13.43	15.50
	50	4.53	5.99	7.36	8.70	10.05	11.43	12.84
	$\infty(\overline{\chi}^2)$	4.23	5.44	6.50	7.48	8.41	9.29	10.16
0.01	10	12.85	19.50	29.30	45.46	76.18	148.17	391.4
	30	8.64	10.95	13.24	15.62	18.15	20.87	23.85
	50	8.05	9.97	11.76	13.53	15.30	17.10	18.95
	$\infty(\overline{\chi}^2)$	7.29	8.75	10.02	11.20	12.26	13.30	14.30

Table 2: Simulation Estimates of the Type I Error Probability in the Homoscedastic Case

α	ν	Approx.	m 2	4	6	8
0.10	10	$\overline{\chi}^2$	0.1478	0.2557	0.4307	0.6987
		$\overline{F}$	0.0983	0.1018	0.0970	0.1041
	30	$\overline{\chi}^2$	0.1188	0.1380	0.1761	0.2252
		$\overline{F}$	0.1029	0.0995	0.0967	0.0961
	50	$\overline{\chi}^2$	0.1027	0.1245	0.1452	0.1620
		$\overline{F}$	0.0948	0.0975	0.1004	0.0961
0.05	10	$\overline{\chi}^2$	0.0940	0.1835	0.3550	0.6285
		$\overline{F}$	0.0497	0.0449	0.0464	0.0505
	30	$\overline{\chi}^2$	0.0646	0.0832	0.1066	0.1507
		$\overline{F}$	0.0513	0.0482	0.0476	0.0447
	50	$\overline{\chi}^2$	0.0588	0.0660	0.0830	0.0997
		$\overline{F}$	0.0507	0.0479	0.0510	0.0464
0.01	10	$\overline{\chi}^2$	0.0348	0.0946	0.2332	0.5149
		$\overline{F}$	0.0106	0.0110	0.0088	0.0091
	30	$\overline{\chi}^2$	0.0183	0.0258	0.0392	0.0610
		$\overline{F}$	0.0109	0.0095	0.0098	0.0081
	50	$\overline{\chi}^2$	0.0150	0.0182	0.0228	0.0318
		$\overline{F}$	0.0101	0.0090	0.0096	0.0074

Table 3: Simulation Estimates of the Type I Error Probability in the Heteroscedastic Case Using the $\overline{F}$ Approximation with Estimated Degrees of Freedom ($\alpha = .05$)

σ_1^2	$\sigma_1'^2$	σ_2^2	ρ_1	ρ_2	$n_1 = n_2 = 20$		$n_1 = n_2 = 30$		$n_1 = n_2 = 50$	
					$m = 4$	$m = 8$	$m = 4$	$m = 8$	$m = 4$	$m = 8$
4	4	1	0	0	.0535	.0562*	.0536	.0486	.0506	.0487
4	2	1	0	0	.0463	.0509	.0530	.0468	.0501	.0500
4	4	1	0.5	0	.0478	.0512	.0481	.0453	.0486	.0500
4	2	1	0.5	0	.0493	.0444	.0466	.0460	.0512	.0460
4	4	1	0	0.5	.0500	.0575*	.0500	.0501	.0458	.0477
4	2	1	0	0.5	.0479	.0546*	.0479	.0474	.0449	.0476
4	4	1	0.5	0.5	.0572*	.0538	.0469	.0525	.0479	.0482
4	2	1	0.5	0.5	.0509	.0463	.0499	.0450	.0473	.0490

* These type I error probability estimates exceed the nominal $\alpha = .05$ at the 5% significance level.

Table 4: Log-Transformed Interleukin-6 Plasma Concentration Measurements at Five Successive Occasions on Bypass Surgery Patients with and without Autotransfusion

Patient	With Autotransfusion					Without Autotransfusion				
j	x_{1j1}	x_{1j2}	x_{1j3}	x_{1j4}	x_{1j5}	x_{2j1}	x_{2j2}	x_{2j3}	x_{2j4}	x_{2j5}
1	2.84	4.00	3.45	2.55	2.46	2.60	3.76	2.86	2.41	2.71
2	2.51	3.26	3.10	2.82	2.48	2.82	3.66	3.20	2.49	2.49
3	2.41	4.14	3.37	2.99	3.04	2.18	3.65	3.87	3.00	2.65
4	2.95	3.42	2.82	3.37	3.35	3.46	3.60	2.97	1.80	1.74
5	3.14	3.25	3.31	2.87	3.41	4.01	3.48	4.42	3.06	2.76
6	3.79	4.34	3.88	3.40	3.16	3.04	2.87	2.87	2.71	2.87
7	4.14	4.97	4.25	3.43	3.06	3.47	3.24	3.47	3.26	3.14
8	3.85	4.31	3.92	3.58	3.91	4.06	3.92	3.18	3.06	1.74
9	3.02	3.11	2.20	2.24	2.28	2.91	3.99	3.06	2.02	3.18
10	3.45	3.41	3.80	3.86	3.91	3.59	4.21	4.02	3.26	2.85
11	5.37	5.02	4.59	3.99	4.27	4.51	4.21	3.78	2.63	1.92
12	3.81	4.21	4.08	3.18	1.86	3.16	3.31	3.28	3.25	3.52
13	4.19	4.59	4.79	4.17	2.60	3.86	3.61	3.28	3.19	3.09
14	3.16	5.30	4.69	4.83	4.51	3.31	2.97	3.76	3.18	2.60
15	3.84	4.32	4.25	3.87	2.93	3.02	2.73	3.87	3.50	2.93