STATISTICS IN MEDICINE
Statist. Med. 2004; **23**:2975–2988 (DOI: 10.1002/sim.1883)

Multiple comparisons with a control in families with both one-sided and two-sided hypotheses

Siu Hung Cheung[1,*,†], Koon Shing Kwong[2], Wai Sum Chan[3] and Shun Piu Leung[4]

[1]*Department of Statistics, Chinese University of Hong Kong, Hong Kong, China*
[2]*Department of Statistics and Applied Probability, National University of Singapore, Singapore*
[3]*Department of Statistics and Actuarial Science, University of Hong Kong, Hong Kong, China*
[4]*Department of Clinical Oncology, Chinese University of Hong Kong, Hong Kong, China*

SUMMARY

Comparing several treatments with a control is a common objective of clinical studies. However, existing procedures mainly deal with particular families of inferences in which all hypotheses are either one- or two-sided. In this article, we seek to develop a procedure which copes with a more general testing environment in which the family of inferences is composed of a mixture of one- and two-sided hypotheses. The proposed procedure provides a more flexible and powerful tool than the existing method. The superiority of this method is also substantiated by a simulation study of average power. Selected critical values are tabulated for the implementation of the proposed procedure. Finally, we provide an illustrative example with sample data extracted from a medical experiment. Copyright © 2004 John Wiley & Sons, Ltd.

KEY WORDS: Dunnett procedure; multiple comparisons with a control; directional-mixed families; familywise type I error

1. INTRODUCTION

In clinical studies, two-sided tests are very common. However, when appropriate, one-sided tests should be adopted to gain greater power [1–3]. For instance, performing one-sided tests to compare a drug to the placebo is likely to be proper in many clinical studies. As explained by Overall [1]:

> ...the sole interest of the sponsor is to document superiority of drug over placebo, and there is no possibility that a sponsor will approach the FDA with a claim of

*Correspondence to: Siu Hung Cheung, Department of Statistics, Chinese University of Hong Kong, Shatin, N.T., Hong Kong, China.
†E-mail: b1@sparc2.sta.cuhk.edu.hk

Contract/grant sponsor: Research Grants Council (Direct Grant) of the Hong Kong Special Administrative Region, China; contract/grant number: 2060190

Received June 2003
Accepted April 2004

> placebo superiority....A one-sided test of significance ensures that the probability of an erroneous conclusion concerning superiority of drug over placebo is less than or equal to the specified alpha level. A two-sided test seriously overestimates the actual probability that chance results will be the basis for an erroneous claim that drug is superior to placebo.

In general, the objectives of a study often dictate the posture of its design [3]. Given the requirements of the regulatory agencies, Koch [3] advised that one-sided tests might be adequate for the following possible settings:

1. a study to demonstrate that a test drug has better efficacy than a placebo;
2. a study to demonstrate that a combination drug has better efficacy than any of its components;
3. a study to demonstrate that a tolerance is an upper bound for the extent to which a test drug has poorer efficacy than an active reference control drug; or
4. a study to demonstrate that a tolerance is an upper bound for the extent to which a test drug has poorer safety than a reference control drug.

As agreed upon by many researchers [1–3], the determination of whether a test is one-sided or two-sided must not be *post hoc*. One-sided inferential posture should be clearly stated in the design protocol. The decision is governed by a number of possible factors such as the *prior* knowledge of the efficacy of the drugs (whether a drug is well-known to be at least as good as another drug) and the intended objective of the drug sponsor. There is no fixed demarcation rule guiding the use of one-sided tests. Rather, the adoption of one-sided tests is a complex decision which relies on many situational factors in particular circumstances.

In clinical studies where the Dunnett's procedure is applied ('one vs many' comparisons), all hypotheses in the family of inferences are usually tested as two-sided by employing the Dunnett's two-sided test [4] (DT2). However, if some hypotheses in the inferential family can be justified to be tested one-sided, the negligence of this information is likely to cause a substantial drop in the overall power of the tests.

To illustrate, let us review a clinical study that was reported by Schwartz *et al.* [5]. The study compared the renal effects of rofecoxib and celecoxib to naproxen in elderly subjects receiving a normal-salt diet. Each elderly subject received one of the four different treatments:

A. rofecoxib;
B. celecoxib;
C. naproxen (active control); or
P. no treatment (the placebo control).

The response variable is the change from baseline for daily urinary sodium excretion during the first 72 h of treatment. Naproxen is a non-steroidal anti-inflammatory drug (NSAID) which is extensively used by elderly subjects [5]. Even though it is associated with therapeutic benefits, it also produces adverse renal effects [6–8]. In recent years, cyclooxygenzse-2 selective inhibitors (e.g. rofecoxib and celecoxib) have been introduced as alternatives to traditional NSAIDs [5]. The experiment is to explore whether rofecoxib and celecoxib produce different renal effects when compared with naproxen.

As the main objective of the experiment is to compare the renal effects of rofecoxib and celecoxib to that of naproxen (active control), the family of inferences contains the tests of A

vs C and B vs C. In addition, as insisted by D′agostino and Heeren [9], in a well-controlled clinical study, the testing for 'downside' sensitivity comparing C to P must be included in order to have a valid trial. Nevertheless, comparisons of A and B to P are not crucial related to the objective of this study.

Since the renal effects of naproxen has a long research history [6–8] and the main purpose of the inclusion of the placebo is to ensure a valid study, a one-sided test of C vs P is suitable. However, there is no compelling reason for one-sided tests between A and C, or B and C. Therefore, the family of inferences consists 2 two-sided tests and 1 one-sided test in this clinical study.

The above example provides the motivation to develop a powerful testing strategy to ascertain how the directional information of some of the comparisons can be effectively incorporated into a 'one vs many' comparison environment. Our findings demonstrate the benefits of the proposed method, which yields a significant improvement in the power of the tests when compared with DT2.

Section 2 presents the proposed procedure. Selected critical values are tabulated in Section 3 while the derivation and computation of the required critical values for the testing procedure are discussed in the appendix. To demonstrate the superiority of our procedure to DT2, Section 4 provides a simulation study. Finally, in Section 5, the testing algorithm is illustrated with sample data extracted from the clinical study of the renal effects of several drugs in elderly subjects by Schwartz *et al.* [5].

2. THE TESTING PROCEDURE

Consider a one-way fixed effect model with $m+1$ treatments,

$$Y_{ij} = \mu_i + \varepsilon_{ij}, \quad i = 0, 1, \ldots, m, \;\; j = 1, \ldots, n_i$$

where Y_{ij} represents the jth observation on the ith treatment, μ_i is the ith treatment mean, and ε_{ij} is a random error component. The sample sizes of the m treatments and the control are n_i (for $i = 1, \ldots, m$) and n_0, respectively, and 0 subscript denotes the control. Hence, the total sample size, denoted by N, is $\sum_{i=0}^{m} n_i$. Assume that $\varepsilon_{ij} \overset{\text{ind}}{\sim} \mathrm{N}(0, \sigma^2)$, where σ^2 is the unknown common variance. Let $\bar{Y}_i$ be the sample mean of ith treatment and $\hat{\sigma}^2$ be the pooled sample variance, which is an unbiased estimator of σ^2 and also independent of $\bar{Y}_i$. To compare the m treatments with the control, the m simultaneous inferences of mean differences are: $\mu_0 - \mu_i$ for all $i = 1, \ldots, m$. Our major objective is to develop a single-step procedure for multiple comparisons with a control when both one- and two-sided hypotheses are present in the given inferential family. We therefore seek to conduct the simultaneous testings of the m null hypotheses:

$$H_i\colon \; \mu_0 = \mu_i$$

for $i = 1, \ldots, m$ versus r one-sided alternative hypotheses

$$H_i'\colon \; \mu_0 < \mu_i$$

for $i=1,\ldots,r$ and $(m-r)$ two-sided alternative hypotheses

$$H_j'\colon\ \mu_0 \neq \mu_j$$

for $j=r+1,\ldots,m$. With respect to the family of null hypotheses $\{H_1,\ldots,H_m\}$, a subset $\{H_1,\ldots,H_r\}$ $(r\leqslant m)$ is tested against one-sided alternatives, while the remaining null hypotheses $\{H_{r+1},\ldots,H_m\}$ are tested against two-sided alternatives. Such families are hereafter referred to as the directional-mixed families.

To test the m null hypotheses simultaneously in the directional-mixed family, the test statistics are $T_1,\ldots,T_r,|T_{r+1}|,\ldots,|T_m|$, where

$$T_i=\frac{\bar{Y}_0-\bar{Y}_i}{\hat{\sigma}\sqrt{1/n_i+1/n_0}}$$

for $i=1,\ldots,m$. The variates $T_1,T_2,\ldots,T_m$ have a multivariate t-distribution with f degrees of freedom ($f=N-m-1$) and covariance matrix $\mathbf{\Sigma}=\{\rho_{ij}\}$, for $i,j=1,\ldots,m$, with a product-correlation structure such that

$$\rho_{ij}=\begin{cases} b_i b_j, & i\neq j \\ 1, & i=j \end{cases}$$

where

$$b_i=\sqrt{\frac{n_i}{n_0+n_i}} \tag{1}$$

To control the familywise error rate strongly at α, we compute positive critical constants $c_{1,\alpha}$ and $c_{2,\alpha}$, which are required for one- and two-sided inferences, respectively, in the directional-mixed family such that

$$G(c_{1,\alpha},c_{2,\alpha})=1-\alpha \tag{2}$$

where

$$G(c_{1,\alpha},c_{2,\alpha})=P\{T_i\geqslant -c_{1,\alpha},\ i=1,\ldots,r;\ |T_j|\leqslant c_{2,\alpha},\ j=r+1,\ldots,m\}$$

For a given α, each null hypothesis H_i for $i=1,\ldots,r$ is rejected if the corresponding $T_i<-c_{1,\alpha}$. Similarly, each null hypothesis H_j for $j=r+1,\ldots,m$ is rejected if the corresponding $|T_j|>c_{2,\alpha}$. By inversion, the corresponding $(1-\alpha)$ joint confidence intervals for $\mu_0-\mu_i$ are

$$\left(-\infty,\ \bar{Y}_0-\bar{Y}_i+c_{1,\alpha}\hat{\sigma}\sqrt{1/n_i+1/n_0}\right) \tag{3}$$

where $i=1,\ldots,r$ and for $\mu_0-\mu_j$ are

$$\left(\bar{Y}_0-\bar{Y}_j-c_{2,\alpha}\hat{\sigma}\sqrt{1/n_j+1/n_0},\ \bar{Y}_0-\bar{Y}_j+c_{2,\alpha}\hat{\sigma}\sqrt{1/n_j+1/n_0}\right) \tag{4}$$

where $j=r+1,\ldots,m$. However, for the cases of the one-sided hypotheses are $H_i'\colon\ \mu_0>\mu_i$ for $i=1,\ldots,r$, each null hypothesis H_i is rejected if the corresponding $T_i>c_{1,\alpha}$. Moreover, the corresponding one-sided confidence intervals should be revised to $\left(\bar{Y}_0-\bar{Y}_i-c_{1,\alpha}\hat{\sigma}\sqrt{1/n_i+1/n_0},\infty\right)$.

Note that when $r=0$ and m, the proposed procedure is reduced to the usual two- and one-sided Dunnett procedures, respectively. Therefore, the corresponding values of $c_{1,\alpha}$ for $r=m$ and $c_{2,\alpha}$ for $r=0$ are the respective critical values required for the usual Dunnett procedures, which are denoted by $D_{1,m,\alpha}$ and $D_{2,m,\alpha}$, respectively.

3. TABULATION OF OPTIMAL VALUES $(c^*_{1,\alpha}, c^*_{2,\alpha})$

The derivation and computation of the optimal values $(c^*_{1,\alpha}, c^*_{2,\alpha})$ under a specified selection criterion are discussed in the appendix. Selected optimal values $(c^*_{1,\alpha}, c^*_{2,\alpha})$ are tabulated in Tables I and II for $\alpha=0.05$ and different values of m, r, f, and ρ. A constant ρ stands for equal correlation, and when $\rho=0.5$, $n_0=n_1=\cdots=n_m$. For the case $r=m$ or 0, the optimal values computed with our subroutine match with $D_{1,m,\alpha}$ and $D_{2,m,\alpha}$, respectively, as tabulated by Bechhofer and Dunnett [10]. All calculations for the accompanying tables of critical values were performed on a SUN Enterprise 4000 server. The Fortran coded subroutine is available from the first author.

In Tables I and II, all critical values are smaller than the corresponding $D_{2,m,\alpha}$. This demonstrates that our procedure is more powerful than DT2, yielding higher probability to reject the null hypotheses. As expected, the gain in power is more substantial when r $(<m)$ increases.

Linear interpolation is suggested for the values of f and ρ that are not tabulated. For the case of unequal correlations ρ_{ij}, instead of computing the optimal points by evaluating the integral, one can use the interpolation method suggested by Dunnett [11] as follows. For $1\leqslant i<j\leqslant m$, compute the arithmetic average value $\bar{\rho}$ of the $m(m-1)/2$ correlation coefficients ρ_{ij}. That is

$$\bar{\rho}=\frac{2}{m(m-1)}\sum_{1\leqslant i<j\leqslant m}\rho_{ij}$$

Then, simultaneous linear interpolation that targets both the correlation $\bar{\rho}$ and degrees of freedom f is used to derive approximate optimal points $c^*_{1,\alpha}$ and $c^*_{2,\alpha}$. Denote the approximate optimal values be $\hat{c}^*_{1,\alpha}$ and $\hat{c}^*_{2,\alpha}$.

To understand the effect of disparities of sample sizes on the accuracy of the approximation method, we first define the average values of b_i, $(i=1,\ldots,r)$ and b_j, $(j=r+1,\ldots,m)$ as

$$\bar{b}_1=\frac{1}{r}\sum_{i=1}^{r}b_i$$

and

$$\bar{b}_2=\frac{1}{m-r}\sum_{j=r+1}^{m}b_j$$

respectively. Note that $\bar{b}_1$ and $\bar{b}_2$ are functions of sample sizes related to one- and two-sided testings, respectively.

Numerical studies indicate that the approximation method is quite accurate for practical purposes as far as the difference between $\bar{b}_1$ and $\bar{b}_2$ is not very large. Our recommendation is that if $0.9<\bar{b}_1/\bar{b}_2<1.1$, the approximation method can be used with satisfactory results. For illustrative purpose, the approximation and exact values of the critical values for selected

Table I. Optimal values of $c^*_{1,\alpha}$ and $c^*_{2,\alpha}$ for $m =$ number of treatments excluding control $= 2$–5.

			$c^*_{1,\alpha}$					$c^*_{2,\alpha}$				
m	r	f	$\rho = 0.1$	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9
2	1	20	2.2659	2.2529	2.2313	2.1938	2.1304	2.2792	2.2754	2.2640	2.2431	2.1967
	1	40	2.1860	2.1792	2.1587	2.1228	2.0615	2.2018	2.1960	2.1863	2.1680	2.1261
	1	60	2.1649	2.1547	2.1338	2.0994	2.0417	2.1747	2.1710	2.1622	2.1443	2.1024
	1	100	2.1462	2.1362	2.1157	2.0814	2.0243	2.1546	2.1512	2.1428	2.1256	2.0848
	1	∞	2.1176	2.1064	2.0880	2.0566	2.0004	2.1255	2.1230	2.1144	2.0971	2.0578
3	1	20	2.5026	2.4867	2.4504	2.3846	2.2714	2.5114	2.4982	2.4697	2.4178	2.3062
	1	40	2.4042	2.3887	2.3587	2.3026	2.1963	2.4131	2.4022	2.3769	2.3296	2.2275
	1	60	2.4078	2.3617	2.3310	2.2758	2.1712	2.3733	2.3705	2.3469	2.3016	2.2026
	1	100	2.3483	2.3369	2.3076	2.2529	2.1515	2.3571	2.3466	2.3238	2.2802	2.1832
	1	∞	2.3165	2.3007	2.2755	2.2245	2.1226	2.3197	2.3112	2.2892	2.2473	2.1544
	2	20	2.3976	2.3744	2.3345	2.2712	2.1636	2.4126	2.4066	2.3895	2.3557	2.2819
	2	40	2.3098	2.2903	2.2537	2.1956	2.0944	2.3234	2.3177	2.3036	2.2732	2.2056
	2	60	2.2816	2.2639	2.2280	2.1713	2.0721	2.2949	2.2886	2.2759	2.2470	2.1814
	2	100	2.2603	2.2423	2.2098	2.1523	2.0549	2.2719	2.2670	2.2525	2.2265	2.1621
	2	∞	2.2287	2.2112	2.1792	2.1238	2.0289	2.2380	2.2341	2.2215	2.1965	2.1343
4	1	20	2.6383	2.6301	2.5840	2.4968	2.3514	2.6633	2.6420	2.6011	2.5280	2.3751
	1	40	2.5401	2.5234	2.4828	2.4125	2.2716	2.5484	2.5322	2.4971	2.4308	2.2920
	1	60	2.5124	2.4877	2.4493	2.3805	2.2425	2.5111	2.4974	2.4641	2.4007	2.2660
	1	100	2.4565	2.4647	2.4240	2.3591	2.2235	2.4877	2.4695	2.4382	2.3765	2.2451
	1	∞	2.4360	2.4226	2.3854	2.3229	2.1938	2.4426	2.4295	2.4003	2.3417	2.2145
	2	20	2.5843	2.5545	2.5070	2.4272	2.2852	2.5925	2.5793	2.5452	2.4831	2.3524
	2	40	2.4773	2.4561	2.4144	2.3405	2.2086	2.4879	2.4747	2.4451	2.3906	2.2710
	2	60	2.4460	2.4233	2.3819	2.3118	2.1843	2.4530	2.4420	2.4147	2.3616	2.2448
	2	100	2.4181	2.3988	2.3584	2.2907	2.1643	2.4274	2.4158	2.3899	2.3383	2.2250
	2	∞	2.3800	2.3612	2.3241	2.2590	2.1363	2.3881	2.3780	2.3532	2.3041	2.1947
	3	20	2.4982	2.4663	2.4141	2.3310	2.1886	2.5135	2.5072	2.4852	2.4433	2.3536
	3	40	2.4023	2.3755	2.3292	2.2514	2.1181	2.4153	2.4096	2.3904	2.3557	2.2732
	3	60	2.3721	2.3457	2.3017	2.2267	2.0955	2.3831	2.3794	2.3607	2.3264	2.2474
	3	100	2.3473	2.3234	2.2802	2.2063	2.0777	2.3597	2.3544	2.3374	2.3053	2.2276
	3	∞	2.3126	2.2898	2.2456	2.1777	2.0519	2.3223	2.3183	2.3073	2.2717	2.1973
5	1	20	2.7747	2.7352	2.6837	2.5915	2.4078	2.7690	2.7468	2.6959	2.6051	2.4241
	1	40	2.6242	2.6197	2.5705	2.4906	2.3209	2.6489	2.6271	2.5842	2.5036	2.3383
	1	60	2.6027	2.5826	2.5350	2.4584	2.2939	2.6071	2.5889	2.5486	2.4713	2.3107
	1	100	2.5684	2.5539	2.5074	2.4300	2.2735	2.5767	2.5590	2.5207	2.4464	2.2892
	1	∞	2.4728	2.5143	2.4636	2.3954	2.2416	2.5380	2.5147	2.4800	2.4088	2.2575
	2	20	2.7132	2.6803	2.6206	2.5318	2.3593	2.7187	2.6990	2.6551	2.5719	2.4059
	2	40	2.5954	2.5687	2.5154	2.4359	2.2797	2.6007	2.5841	2.5469	2.4731	2.3203
	2	60	2.5726	2.5337	2.4857	2.4051	2.2519	2.5581	2.5472	2.5112	2.4417	2.2937
	2	100	2.5278	2.5058	2.4591	2.3822	2.2318	2.5336	2.5185	2.4844	2.4167	2.2725
	2	∞	2.4631	2.4634	2.4213	2.3500	2.2012	2.4975	2.4767	2.4444	2.3794	2.2413
	3	20	2.6454	2.6168	2.5562	2.4624	2.2985	2.6651	2.6450	2.6089	2.5393	2.3934
	3	40	2.5340	2.5106	2.4584	2.3728	2.2202	2.5534	2.5363	2.5045	2.4432	2.3101

Table I. *Continued*

			$c^*_{1,\alpha}$					$c^*_{2,\alpha}$				
m	r	f	$\rho=0.1$	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9
	3	60	2.5045	2.4772	2.4277	2.3441	2.1961	2.5131	2.5011	2.4709	2.4126	2.2827
	3	100	2.4802	2.4511	2.4033	2.3219	2.1762	2.4825	2.4736	2.4448	2.3886	2.2622
	3	∞	2.4354	2.4131	2.3662	2.2889	2.1473	2.4440	2.4328	2.4073	2.3532	2.2313
	4	20	2.5795	2.5409	2.4779	2.3793	2.2085	2.5939	2.5873	2.5629	2.5152	2.4168
	4	40	2.4754	2.4440	2.3884	2.2974	2.1375	2.4905	2.4831	2.4624	2.4216	2.3317
	4	60	2.4412	2.4130	2.3595	2.2712	2.1147	2.4591	2.4499	2.4308	2.3918	2.3047
	4	100	2.4175	2.3885	2.3372	2.2511	2.0969	2.4284	2.4244	2.4054	2.3678	2.2836
	4	∞	2.3789	2.3530	2.3039	2.2208	2.0704	2.3904	2.3853	2.3688	2.3334	2.2526

Table II. Optimal values of $c^*_{1,\alpha}$ and $c^*_{2,\alpha}$ for $m=$ number of treatments excluding control $=6,\ 7$.

			$c^*_{1,\alpha}$					$c^*_{2,\alpha}$				
m	r	f	$\rho=0.1$	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9
6	1	20	2.8265	2.8179	2.7544	2.6512	2.4446	2.8592	2.8291	2.7707	2.6669	2.4625
	1	40	2.7357	2.6974	2.6452	2.5430	2.3607	2.7225	2.7011	2.6514	2.5608	2.3735
	1	60	2.6982	2.6519	2.5987	2.5164	2.3330	2.6797	2.6610	2.6146	2.5263	2.3452
	1	100	2.6222	2.6242	2.5762	2.4921	2.3109	2.6508	2.6288	2.5844	2.4996	2.3231
	1	∞	2.5740	2.5774	2.5311	2.4545	2.2758	2.6020	2.5821	2.5412	2.4605	2.2907
	2	20	2.8136	2.7757	2.7120	2.6071	2.4152	2.8147	2.7905	2.7366	2.6400	2.4461
	2	40	2.6701	2.6539	2.5996	2.5056	2.3270	2.6907	2.6671	2.6211	2.5358	2.3593
	2	60	2.6399	2.6174	2.5635	2.4735	2.3003	2.6470	2.6271	2.5843	2.5026	2.3312
	2	100	2.5911	2.5881	2.5354	2.4486	2.2773	2.6194	2.5960	2.5556	2.4765	2.3098
	2	∞	2.6407	2.5406	2.4953	2.4111	2.2473	2.5501	2.5513	2.5130	2.4383	2.2771
	3	20	2.7667	2.7245	2.6593	2.5566	2.3685	2.7681	2.7490	2.7007	2.6120	2.4337
	3	40	2.6421	2.6088	2.5515	2.4580	2.2861	2.6461	2.6296	2.5888	2.5109	2.3471
	3	60	2.6051	2.5722	2.5176	2.4270	2.2587	2.6057	2.5913	2.5529	2.4785	2.3200
	3	100	2.5690	2.5439	2.4904	2.4027	2.2391	2.5776	2.5612	2.5252	2.4532	2.2980
	3	∞	2.5112	2.5021	2.4522	2.3672	2.2082	2.5388	2.5169	2.4834	2.4157	2.2662
	4	20	2.7081	2.6668	2.5996	2.4927	2.3086	2.7197	2.7040	2.6633	2.5888	2.4321
	4	40	2.5927	2.5573	2.4983	2.4015	2.2307	2.6012	2.5895	2.5549	2.4889	2.3456
	4	60	2.5556	2.5238	2.4664	2.3730	2.2061	2.5636	2.5515	2.5199	2.4566	2.3177
	4	100	2.5290	2.4959	2.4415	2.3499	2.1864	2.5317	2.5234	2.4925	2.4320	2.2963
	4	∞	2.4824	2.4554	2.4043	2.3160	2.1564	2.4918	2.4814	2.4526	2.3956	2.2656
	5	20	2.6448	2.6020	2.5313	2.4197	2.2258	2.6669	2.6564	2.6280	2.5761	2.4722
	5	40	2.5365	2.5008	2.4382	2.3358	2.1538	2.5529	2.5451	2.5217	2.4774	2.3844
	5	60	2.5019	2.4683	2.4083	2.3090	2.1307	2.5163	2.5098	2.4881	2.4458	2.3562
	5	100	2.4744	2.4426	2.3846	2.2880	2.1130	2.4885	2.4826	2.4626	2.4212	2.3336
	5	∞	2.4342	2.4053	2.3504	2.2567	2.0862	2.4468	2.4414	2.4230	2.3856	2.3015
7	1	20	2.8979	2.8799	2.8291	2.7113	2.4801	2.9300	2.8972	2.8303	2.7160	2.4929
	1	40	2.8160	2.7506	2.7008	2.5931	2.3936	2.7850	2.7627	2.7068	2.6065	2.4019
	1	60	2.7439	2.7240	2.6590	2.5596	2.3637	2.7424	2.7186	2.6676	2.5712	2.3731

Table II. *Continued*

			$c^*_{1,\alpha}$					$c^*_{2,\alpha}$				
m	r	f	$\rho=0.1$	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9
	1	100	2.7389	2.6809	2.6293	2.5386	2.3413	2.7049	2.6859	2.6366	2.5433	2.3505
	1	∞	2.6806	2.6342	2.5786	2.4912	2.3077	2.6537	2.6364	2.5917	2.5034	2.3171
	2	20	2.8859	2.8546	2.7807	2.6657	2.4535	2.8943	2.8640	2.8033	2.6946	2.4796
	2	40	2.7529	2.7225	2.6649	2.5636	2.3642	2.7573	2.7337	2.6813	2.5853	2.3902
	2	60	2.7130	2.6812	2.6265	2.5283	2.3360	2.7130	2.6920	2.6427	2.5510	2.3615
	2	100	2.6624	2.6540	2.5984	2.5007	2.3148	2.6822	2.6583	2.6122	2.5243	2.3390
	2	∞	2.7238	2.5988	2.5532	2.4661	2.2837	2.6123	2.6118	2.5680	2.4836	2.3056
	3	20	2.8563	2.8071	2.7377	2.6240	2.4181	2.8542	2.8314	2.7739	2.6719	2.4681
	3	40	2.7187	2.6852	2.6237	2.5211	2.3314	2.7242	2.7032	2.6551	2.5655	2.3795
	3	60	2.6796	2.6457	2.5894	2.4909	2.3053	2.6808	2.6624	2.6165	2.5306	2.3507
	3	100	2.6369	2.6147	2.5598	2.4648	2.2845	2.6514	2.6307	2.5873	2.5043	2.3283
	3	∞	2.5574	2.5687	2.5182	2.4253	2.2523	2.6153	2.5840	2.5437	2.4659	2.2957
	4	20	2.8004	2.7636	2.6908	2.5752	2.3753	2.8200	2.7936	2.7428	2.6506	2.4604
	4	40	2.6773	2.6437	2.5811	2.4781	2.2923	2.6900	2.6707	2.6272	2.5452	2.3724
	4	60	2.6369	2.6066	2.5469	2.4469	2.2659	2.6491	2.6307	2.5898	2.5118	2.3442
	4	100	2.6028	2.5769	2.5198	2.4219	2.2450	2.6188	2.5999	2.5608	2.4861	2.3223
	4	∞	2.5625	2.5333	2.4798	2.3862	2.2149	2.5675	2.5543	2.5182	2.4474	2.2895
	5	20	2.7559	2.7114	2.6357	2.5200	2.3185	2.7739	2.7554	2.7135	2.6324	2.4670
	5	40	2.6368	2.5994	2.5335	2.4272	2.2392	2.6489	2.6349	2.5991	2.5293	2.3794
	5	60	2.5981	2.5631	2.5003	2.3980	2.2147	2.6097	2.5970	2.5633	2.4959	2.3504
	5	100	2.5676	2.5354	2.4749	2.3744	2.1951	2.5793	2.5665	2.5346	2.4707	2.3283
	5	∞	2.5239	2.4936	2.4370	2.3398	2.1659	2.5328	2.5227	2.4929	2.4334	2.2958
	6	20	2.7031	2.6551	2.5772	2.4541	2.2396	2.7244	2.7143	2.6833	2.6295	2.5256
	6	40	2.5884	2.5495	2.4807	2.3687	2.1679	2.6074	2.5978	2.5730	2.5254	2.4318
	6	60	2.5528	2.5155	2.4497	2.3412	2.1447	2.5665	2.5612	2.5381	2.4927	2.4026
	6	100	2.5245	2.4892	2.4255	2.3197	2.1267	2.5358	2.5314	2.5109	2.4671	2.3796
	6	∞	2.4817	2.4500	2.3899	2.2879	2.1001	2.4936	2.4890	2.4699	2.4295	2.3448

configurations of sample sizes with $m=4$ are provided in Table III. Notice that even the disparities of sample sizes are drastic such as Case A ($r=2$), the approximation is excellent because $\bar{b}_1/\bar{b}_2=1.05$ which is close to 1.

4. POWER STUDY

To evaluate the gain in power as compared to DT2 in a directional-mixed family, a simulation study is performed to study the average power (the proportion of false hypotheses that are correctly rejected). Average power is calculated based on simulations with 100 000 replications. Figures 1(a) and 1(b) present the percentage increase in average power when our procedure is compared to DT2 with $m=3$ and 10 ($f=60$ and $\rho=0.5$). Other choices of f and ρ have been studied and similar patterns obtained, and hence are not reported. The

Table III. Values of $(c^*_{1,\alpha}, c^*_{2,\alpha})$ and $(\hat{c}^*_{1,\alpha}, \hat{c}^*_{2,\alpha})$ for selected configurations of sample sizes and $m = 4$.

Sample sizes	r	$\bar{b}_1/\bar{b}_2$	$c^*_{1,\alpha}$	$c^*_{2,\alpha}$	$\hat{c}^*_{1,\alpha}$	$\hat{c}^*_{2,\alpha}$
Case A	1	1.30	2.467	2.429	2.356	2.437
$(n_0, n_1, n_2, n_3, n_4) = (10, 35, 6, 40, 4)$	2	1.05	2.362	2.378	2.356	2.389
	3	1.49	2.314	2.271	2.276	2.339
Case B	1	1.08	2.429	2.434	2.414	2.431
$(n_0, n_1, n_2, n_3, n_4) = (10, 20, 10, 17, 14)$	2	0.98	2.344	2.389	2.346	2.387
	3	1.01	2.266	2.341	2.264	2.342
Case C	1	0.78	2.402	2.479	2.465	2.479
$(n_0, n_1, n_2, n_3, n_4) = (10, 6, 15, 20, 7)$	2	0.93	2.375	2.431	2.397	2.428
	3	1.13	2.321	2.352	2.339	2.373

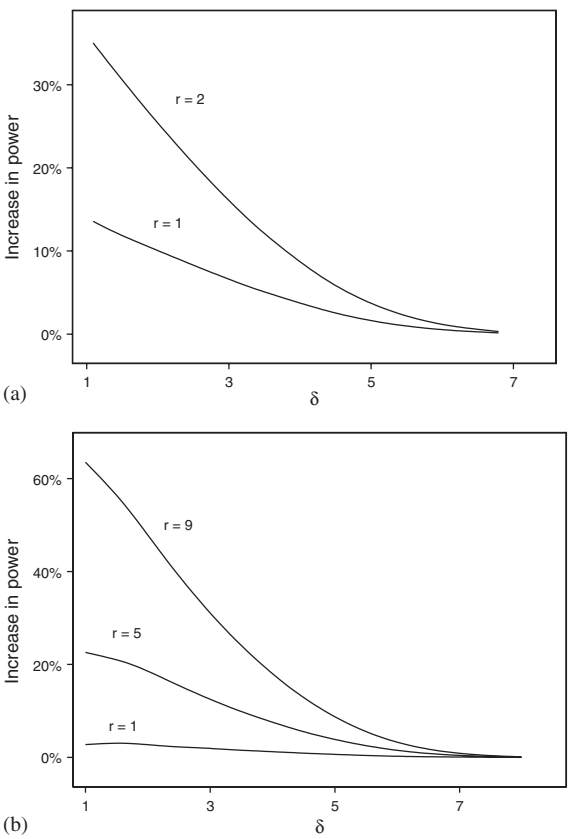

Figure 1. (a) Increase in average power as compared to DT2 when $m = 3$; and (b) increase in average power as compared to DT2 when $m = 10$.

Table IV. Change from baseline for daily urinary sodium excretion in elderly subjects treated with rofecoxib, celecoxib, naproxen, or placebo.

Naproxen ($n_0 = 15$)	Placebo ($n_1 = 14$)	Rofecoxib ($n_2 = 17$)	Celecoxib ($n_3 = 16$)
−40.6	−10.8	−39.5	−27.1

Daily urinary sodium excretion is measured in milliequivalents per 24 h during the first 72 h, $\hat{\sigma} = 25$ with $f = 58$.

non-centrality parameter of $\bar{Y}_i$ for $i = 1, \ldots, m$ is $\delta\sigma/\sqrt{n}$. The zero mean is used to generate $\bar{Y}_0$. The percentage gain in power is plotted against δ for $\delta = 1$ to $\delta = 8$.

These figures reveal two major characteristics. First, the gain in average power is drastic for large r, which represents the case where there are many one-sided inferences, thus allowing the incorporation of more useful prior information to enhance the power of our testing procedure. Second, the gain in power is ample for smaller δ. As δ becomes larger, it is reasonable to expect that both DT2 and our procedure will be able to reject most of the hypotheses, thus yielding very similar average power for both procedures. Note that the gain in average power can be very dramatic in some cases. For example, when $m = 10$, $r = 9$, and $\delta = 1$, the average power of our procedure is over 60 per cent higher than that of DT2.

5. EXAMPLE

To illustrate the testing procedure, the clinical study discussed in Section 1 is revisited here. There are four groups of elderly subjects receiving four different treatments. The data are shown in Table IV.

As explained in Section 1, naproxen (the active control) is compared with the placebo with an one-sided test while two-sided tests are conducted for the remaining hypotheses in the family. The response variable is the daily urinary sodium excretion, measured in milliequivalents per 24 h during the first 72 h.

Let $\mu_0, \mu_1, \mu_2, \mu_3$ be the true mean changes from baseline for daily urinary sodium excretion of the four groups: naproxen, placebo, rofecoxib and celecoxib, respectively. The simultaneous testing of the three null hypotheses is as follows:

$$H_i\colon\ \mu_0 = \mu_i$$

for $i = 1, \ldots, 3$ versus

$$H_i'\colon\ \mu_0 < \mu_i$$

for $i = 1$ and

$$H_j'\colon\ \mu_0 \neq \mu_j$$

for $j = 2, 3$. In this example, $(n_0, n_1, n_2, n_3) = (15, 14, 17, 16)$, $m = 3$, $r = 1$, $f = 58$. With $\alpha = 0.05$, the exact critical values of the proposed procedure are $c^*_{1,0.05} = 2.318$ and $c^*_{2,0.05} = 2.350$. Since the ratio $\bar{b}_1/\bar{b}_2 = 0.922$, the approximation method may be used. With $\bar{\rho} = 0.509741$, the approximation values are $\hat{c}^*_{1,0.05} = 2.331$ and $\hat{c}^*_{2,0.05} = 2.347$. On the other hand, the critical

value for DT2 is 2.41. Smaller critical values $c^*_{1,0.05}$ and $c^*_{2,0.05}$ as compared to the critical value of DT2 indicate that the proposed procedure is more powerful in this directional-mixed family of inferences. According to (3) and (4), the 95 per cent simultaneous confidence intervals for $\mu_0 - \mu_i$ $(i = 1, 2, 3)$ are:

1. $\mu_0 - \mu_1$: $(-\infty,\ -8.265)$;
2. $\mu_0 - \mu_2$: $(-21.912,\ 19.712)$; and
3. $\mu_0 - \mu_3$: $(-34.615,\ 7.615)$.

It is obvious that only the first null hypothesis is rejected, as zero is not contained in the corresponding confidence interval. Therefore, the data pass the sensitivity test and support the validity of the study. However, the non-rejections of the other two-sided hypotheses provide no evidence to support that rofecoxib and celecoxib (cyclooxygenzse-2 (COX-2) inhibitors) have different renal effects as compared with naproxen. As explained by Schwartz *et al.* [5], based on these findings, 'it is recommended that clinicians, as they would for non-selective NSAIDs, carefully follow the renal precautions for COX-2 selective inhibitors that are present in their respective product circulars'.

APPENDIX A: EVALUATION OF CRITICAL VALUES

For a given α and $r = 1, \ldots, m-1$, there are infinite pairs of possible solutions of $\{c_{1,\alpha}, c_{2,\alpha}\}$ that satisfy equation (2). Hence, we have to propose a criterion for the selection of $\{c_{1,\alpha}, c_{2,\alpha}\}$ in the directional-mixed family of inferences.

For two-sided confidence intervals $(\bar{Y}_0 - \bar{Y}_j \pm \Delta_{2j})$, where $\Delta_{2j} = c_{2,\alpha}\hat{\sigma}\sqrt{1/n_j + 1/n_0}$, $(j = r+1, \ldots, m)$, a minimum value of Δ_{2j} is a desirable property related to the concept of uniformly most accurate (UMA) interval (see for example, Reference [12, Chapter 9]). In essence, a smaller value of Δ_{2j} yields a shorter interval and hence an increase in precision. As the result, Spurrier and Nizam [13] used the UMA idea for the derivation of optimal allocation of sample sizes by minimizing the expected average allowance (EAA), which is defined as $E(2\sum_{j=1}^{m} \Delta_{2j})$ for the case of $r = 0$.

Similarly, for one-sided intervals $(-\infty,\ \bar{Y}_0 - \bar{Y}_i + \Delta_{1i})$ where $\Delta_{1i} = c_{1,\alpha}\hat{\sigma}\sqrt{1/n_i + 1/n_0}$, the idea of UMA requires the minimum of Δ_{1i} which implies a shorter interval. Therefore, EAA is conceptually extended to $E(\sum_{i=1}^{m} \Delta_{1i})$ for the cases of $r = m$.

By generalization the idea of UMA in the directional-mixed family of inferences, i.e. $r = 1, \ldots, m-1$, we modify and redefine EAA as

$$\mathrm{EAA} = d_1 c_{1,\alpha} + d_2 c_{2,\alpha} \tag{A1}$$

where

$$d_1 = \frac{E(\hat{\sigma})}{m} \sum_{i=1}^{r} \sqrt{1/n_i + 1/n_0}$$

and

$$d_2 = \frac{2E(\hat{\sigma})}{m} \sum_{j=r+1}^{m} \sqrt{1/n_j + 1/n_0}$$

Let the optimum critical values of $c_{1,\alpha}$ and $c_{2,\alpha}$ that minimize EAA with a given α level be $c^*_{1,\alpha}$ and $c^*_{2,\alpha}$, subject to the non-linear equation (2). With respect to equation (2) with a given α, $c_{1,\alpha}$ can be written as a decreasing function of $c_{2,\alpha}$, say $c_{1,\alpha}=H(c_{2,\alpha})$. To facilitate computation and reduce the workload of searching for the optimal critical values, we derive the bounds of $c_{1,\alpha}$ and $c_{2,\alpha}$ as follows.

1. Upper bounds of $c_{1,\alpha}$ and $c_{2,\alpha}$.
   To compare several treatments with a control in the directional-mixed families, the common practice is to use DT2. As our objective is to design a procedure which is superior to DT2, it is desirable to impose a property which guarantees that the new procedure dominates DT2. That is, all hypotheses that are rejected by DT2 will also be rejected by the new procedure. To this end, the values of both $c_{1,\alpha}$ and $c_{2,\alpha}$ are restricted to be at most $D_{2,m,\alpha}$.
2. Lower bounds of $c_{1,\alpha}$ and $c_{2,\alpha}$.
   As H is a decreasing function of $c_{2,\alpha}$ and $D_{2,m,\alpha}$ is the upper bound of $c_{2,\alpha}$, the minimum value of $c_{1,\alpha}$, denoted $c_{1,\alpha,\min}$, is equal to $H(D_{2,m,\alpha})$. Similarly, the minimum value of $c_{2,\alpha}$, denoted $c_{2,\alpha,\min}$, is equal to $H^{-1}(D_{2,m,\alpha})$.

With the bounds of $c_{1,\alpha}$ and $c_{2,\alpha}$, the optimum values $c^*_{1,\alpha}$ and $c^*_{2,\alpha}$ are computed as follows:

1. Given m, r, α, and f, start with an initial guess of $c_{1,\alpha}$ (for simplicity, take the mean of $D_{2,m,\alpha}$ and $c_{1,\alpha,\min}$), say $c^0_{1,\alpha}$. Then, compute $c_{2,\alpha}$ by equation (2). Note that

$$
\begin{aligned}
&G(c_{1,\alpha}, c_{2,\alpha})\\
&\quad = P\{T_i \geqslant -c_{1,\alpha},\ i=1,\ldots,r;\ |T_j| \leqslant c_{2,\alpha},\ j=r+1,\ldots,m\}\\
&\quad = P\left\{\frac{\bar{Y}_0-\bar{Y}_i}{\hat{\sigma}\sqrt{1/n_i+1/n_0}} \geqslant -c_{1,\alpha},\ i=1,\ldots,r;\ \frac{|\bar{Y}_0-\bar{Y}_j|}{\hat{\sigma}\sqrt{1/n_j+1/n_0}} \leqslant c_{2,\alpha},\ j=r+1,\ldots,m\right\}\\
&\quad = P\left\{\frac{\bar{Y}_0-\bar{Y}_i}{\sigma\sqrt{1/n_i+1/n_0}} \geqslant -\frac{\hat{\sigma}}{\sigma}c_{1,\alpha},\ i=1,\ldots,r;\ \frac{|\bar{Y}_0-\bar{Y}_j|}{\sigma\sqrt{1/n_j+1/n_0}} \leqslant \frac{\hat{\sigma}}{\sigma}c_{2,\alpha},\ j=r+1,\ldots,m\right\}
\end{aligned}
$$

Let $u=\hat{\sigma}/\sigma$. Conditioning on u, the above expression becomes

$$
\int_0^\infty P\left\{\frac{\bar{Y}_0-\bar{Y}_i}{\sigma\sqrt{1/n_i+1/n_0}} \geqslant -uc_{1,\alpha},\ i=1,\ldots,r;\ \frac{|\bar{Y}_0-\bar{Y}_j|}{\sigma\sqrt{1/n_j+1/n_0}} \leqslant uc_{2,\alpha},\ j=r+1,\ldots,m\right\} g(u)\,\mathrm{d}u \tag{A2}
$$

where $g(\cdot)$ is the p.d.f. of a $\sqrt{\chi^2_f/f}$ random variable. Then, follow the arguments of Hochberg and Tamhane [14, p. 374], conditioning on $y=\sqrt{n_0}\bar{Y}_0/\sigma \sim N(0,1)$, (A2)

becomes

$$\int_0^\infty \int_{-\infty}^\infty \prod_{i=1}^{r} \left[\Phi\left(\frac{b_i y + u c_{1,\alpha}}{\sqrt{1-b_i^2}} \right) \right] \prod_{j=r+1}^{m} \left[\Phi\left(\frac{b_j y + u c_{2,\alpha}}{\sqrt{1-b_j^2}} \right) - \Phi\left(\frac{b_j y - u c_{2,\alpha}}{\sqrt{1-b_j^2}} \right) \right]$$

$$\times \phi(y) g(u) \, \mathrm{d}y \, \mathrm{d}u$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ are the standard normal c.d.f. and p.d.f., respectively. Therefore, the evaluation involves the computation of a two-dimensional integration. One can apply the algorithm of Dunnett [15] or that of Cheung and Holland [16, 17] with slight modification. A Fortran coded subroutine is also available on the STATLIB electronic bulletin board (Internet address: http://lib.stat.cmu.edu/apstat/251).

2. Based on the values of $c_{1,\alpha}$ and $c_{2,\alpha}$ obtained above, compute EAA according to equation (A1). From equation (A1), it is obvious that the different values of $E(\hat{\sigma})$ give rise to the same optimal values of $c_{1,\alpha}$ and $c_{2,\alpha}$. Therefore, we simply let $E(\hat{\sigma})$ be 1.
3. Search for the minimum EAA by repeating the previous two steps with other values of $c_{1,\alpha}$ in the range of $(c_{1,\alpha,\min},\ D_{2,m,\alpha})$. The search is performed by the IMSL subroutine UVMIF. The $c_{1,\alpha}$ and $c_{2,\alpha}$ that minimize EAA will be the optimal critical values.

ACKNOWLEDGEMENTS

The authors are grateful to Professor Charles Dunnett, Professor Burt Holland and two referees for providing several helpful comments which led to a significant improvement of the article.

REFERENCES

1. Overall JE. A comment concerning one-sided tests of significance in new drug applications. *Journal of Biopharmaceutical Statistics* 1991; **1**:157–160.
2. Peace K. One-sided or two-sided *p* values: which most appropriately address the question of drug efficacy? *Journal of Biopharmaceutical Statistics* 1991; **1**:133–138.
3. Koch GG. One-sided and two-sided tests and *p* values. *Journal of Biopharmaceutical Statistics* 1991; **1**: 161–170.
4. Dunnett CW. A multiple comparison procedure for comparing several treatments with a control. *Journal of the American Statistical Association* 1955; **50**:1096–1121.
5. Schwartz JI, Vandormael K, Malice MP, Kalyani RN, Lasseter KC, Holmes GB, Gertz BJ, Gottesdiener KM, Laurenzi M, Redfern KJ, Brune K. Comparison of rofecoxib, celecoxib, and naproxen on renal function in elderly subjects receiving a normal-salt diet. *Clinical Pharmacology and Therapeutics* 2002; **72**:50–61.
6. Clive DM, Stoff JS. Renal syndromes associated with nonsteroidal anti-inflammatory drugs. *New England Journal of Medicine* 1984; **310**:563–573.
7. Murray MD, Brater DC. Renal toxicity of the nonsteroidal anti-inflammatory drugs. *Annual Review of Pharmacology and Toxicology* 1993; **32**:435–465.
8. Swan SK, Rudy DW, Lasseter KC, Ryan CF, Buechel KL, Lambrecht LJ, Pinto MB, Dilzer SC, Obrda O, Sundblad KJ, Gumbs CP, Ebel DL, Quan H, Larson PJ, Schwartz JI, Musliner TA, Gertz BJ, Brater DC, Yao S. Effect of cyclooxygenase-2 inhibition on renal function in elderly persons receiving a low-salt diet: a randomized, controlled trial. *Annals of Internal Medicine* 2000; **133**:1–9.
9. D′agostino RB, Heeren TC. Multiple Comparisons in over-the-counter drug clinical trials with both positive and placebo controls. *Statistics in Medicine* 1991; **10**:1–6.
10. Bechhofer RE, Dunnett CW. *Percentage Points of Multivariate Student t Distribution*: *Selected Tables in Mathematical Statistics*, vol. 11. American Mathematical Society: Providence, 1988.
11. Dunnett CW. Multiple comparisons between several treatments and a specified treatment. In *Lecture Notes in Statistics*: *Linear Statistical Inferences*, Calinski T, Klonechi W (eds), vol. 35. Springer: Berlin, 1985; 39–47.

12. Casella G, Berger, RL. *Statistical Inference*. Brooks and Cole: California; 1990.
13. Spurrier JD, Nizam A. Sample size allocation for simultaneous inference in comparison with control experiments. *Journal of the American Statistical Association* 1990; **85**:181–186.
14. Hochberg Y, Tamhane AC. *Multiple Comparison Procedures*. Wiley: New York; 1987.
15. Dunnett CW. Algorithm AS 251: multivariate normal probability integrals with product correlation structure. *Applied Statistics* 1989; **38**:564–579 [Correction Note: *Applied Statistics* 1993; **42**:709].
16. Cheung SH, Holland B. Extension of Dunnett's multiple comparison procedure to the case of several groups. *Biometrics* 1991; **47**:21–32.
17. Cheung SH, Holland B. Extension of Dunnett's multiple comparison procedure with different sample sizes to the case of several groups. *Computational Statistics and Data Analysis* 1992; **14**:165–182.

