ELSEVIER

Journal of Statistical Planning and Inference 82 (1999) 83–99

journal of statistical planning and inference

www.elsevier.com/locate/jspi

Exact unconditional distributions for dichotomous data in many-to-one comparisons

Hans-Friedrich Koch [a,*], Ludwig A. Hothorn [b]

[a] *Solvay Pharmaceutical BV, Weesp, The Netherlands*

[b] *Dahlienweg 18, D30916 Isernhagen, Germany*

Received 1 January 1997; accepted 1 March 1997

Abstract

The exact conditional distribution of a statistic for the many-to-one design with dichotomous response variables was firstly given by Williams (1988). His approach is expanded to other maximum statistics, and their exact conditional distributions are compared. As an alternative to the conditional approach, the exact unconditional binomial distribution is presented for two maximum statistics. For these distributions algorithms are described to compute the exact unconditional p-values and critical values. As an application of the exact unconditional distribution, a new simultaneous step-down procedure is suggested. This procedure is based on the exact unconditional distribution and is compared to the α-adjustment procedure according to Bonferroni–Holm and Hommel when p-values result from Fisher's exact test and Barnard's test. © 1999 Elsevier Science B.V. All rights reserved.

MSC: 62J15; 62E15

Keywords: Many-to-one design; Binary data; Exact unconditional test; Step-down procedure

1. Introduction

The comparison of several populations with one control group is a common application in bio-medical experiments. In this design the control is often a placebo group and the several populations represent different dose levels of the same investigational drug or different treatment regimens. This experimental setting is either referred to as "many-to-one" or "comparison with a control". In the case of normally distributed data, several multiple comparison procedures based on the multivariate t-distribution have been proposed since Dunnett's work in 1955.

The many-to-one design with dichotomous response variables is commonly encountered in clinical trials and toxicological experiments. The observed data may be the

* Corresponding author. Fax: +49-511-611489.

E-mail address: Hans-Friedrich.Koch@Solvay.com (H. Koch)

0378-3758/99/$ - see front matter © 1999 Elsevier Science B.V. All rights reserved.
PII: S0378-3758(99)00033-6

response to a new medical treatment or quantitatively measured data may be classified to binary data, as response or no response at an a priori defined or data-dependent-point. The many-to-one design commonly consists of k-independent treatment groups and a concurrent control group. Within the ith group a success is observed with the probability π_i $(i=0,\ldots,k)$. Therefore, the k success probabilities π_i of the treatment groups are compared with the success probability π_0 of the control group. The analysis of this setting includes k comparisons of π_i with π_0 and correlation necessarily occurs since each comparison is with the same π_0. Some correction for this multiplicity is required in the analysis (Hochberg and Tamhane, 1987).

For the ith treatment the total number r_i of successes is recorded which are independent binomially distributed with n_i observations and success rate π_i, i.e. $R_i \sim B(n_i, \pi_i)$, π_i the success rate of the ith group, $i=0,\ldots,k$. Let $\hat{\pi}_i$ denote the maximum likelihood estimator r_i/n_i and $\mathrm{Var}(\hat{\pi}_i)=\pi_i(1-\pi_i)/n_i$. Consider hypotheses H_{0i}: $\pi_i \leqslant \pi_0$ versus H_{Ai}: $\pi_i > \pi_0$ for $i=1,\ldots,k$. To test these hypothesis Ahner and Passing (1983) suggested the use of the maximum of the statistics based on $T_i=(\hat{\pi}_i-\hat{\pi}_0)/(\pi_\cdot(1-\pi_\cdot))/(1/n_i+1/n_0)$, where $\pi_\cdot=(r_0+\cdots+r_k)/(n_0+\cdots+n_k)$. The asymptotic joint distribution of $(T_1,\ldots,T_k)$ is a multivariate normal distribution with

$$E(T_i)=0, \quad \mathrm{Var}(T_i)=1 \quad \text{and} \quad \mathrm{Cov}(T_i,T_j)=\left\{\left(\frac{1}{(1+\frac{n_0}{n_i})}\right)\left(\frac{1}{(1+\frac{n_0}{n_j})}\right)\right\}^{-1/2}.$$

A generalization of the multinomial distribution was given by Passing (1984). The author investigated the asymptotic distribution and referred to the possibility of using an exact conditional distribution without giving any results. The exact conditional distribution of the maximum statistic was described firstly by Williams (1988), who showed that the test with this asymptotic distribution is anticonservative when the sample sizes are large. Also Fairweather (1987) investigated the conditional distribution in a similar design. Bristol (1993) applied the multivariate normal distribution to compute the power of three maximum statistics which differ from each other only by the variance estimator.

As an extension of Barnard's test for 2×2 tables we propose to use the exact unconditional distribution of the maximum statistic, which depends on the unknown π under the overall null hypothesis H_0. We eliminate the effect of the nuisance parameter π by maximizing the size of the test over the domain of the nuisance parameter. Suissa and Shuster (1985) applied this approach to the computation of critical constants of the exact unconditional distribution of the statistic for the two-sample design.

Multiple comparison for the many-to-one design with binomially distributed random variables have been addressed only by Piegorsch (1991, 1995) and Chuang-Stein and Tong (1995). Piegorsch (1995) suggested to address the multiplicity arising from the comparison of each treatment to the control by dividing the familywise (overall) type I error by the number of comparisons, that is, the Bonferroni adjustment. The type I familywise error (FWE) gives the probability of rejecting any true null-hypothesis. Furthermore, he considered also the modification of this procedure by Holm (1977), here referred as Bonferroni–Holm. It is well known that both procedures

are conservative especially when the number of comparisons are high. Theoretically, one can apply both procedures when dealing with binary data without any restrictions. Finally, Piegorsch (1995) also investigated the property of the above-defined maximum statistics when they are compared in the Dunnett step-down procedure to the critical constants of the multivariate t-distribution. Chuang-Stein and Tong (1995) stated that at least in case of small sample sizes the performance of Dunnett's procedure in the case of binary response is unknown. Therefore, it is desirable to investigate the possibility of constructing a multiple comparison procedure based on the binomial distribution. As a first solution Chuang-Stein and Tong (1995) simulated exact critical values in the unconditional model for fixed overall success probability π.

In Section 2 we will describe the statistics used and compute the exact conditional as well as the exact unconditional distribution. Results are presented in Section 3. The practical properties of the new method are discussed in Section 4 where it is compared to the existing methods. In Section 5 a new simultaneous multiple step-down procedure based on the exact unconditional distribution is introduced. An example is analysed with the new procedure and the bootstrap method in Section 6. An extension to two-sided tests is given in Section 7.

2. Methods

In accordance with Williams (1988) and Bristol (1993) we investigated the following five statistics. Throughout, the notation $n_\bullet = n_0 + \cdots + n_k$ and $r_\bullet = r_0 + \cdots + r_k$ is used.

1. Berkson's difference statistic T^{D}

$$T^{\mathrm{D}} = \max_{i\in\{1,\ldots,k\}} T_i^{\mathrm{D}}, \quad \text{with } T_i^{\mathrm{D}} = \hat{\pi}_i - \hat{\pi}_0, \quad i \in \{1,\ldots,k\}.$$

2. A statistic with unpooled variance estimation

$$T^{\mathrm{UP}} = \max_{i\in\{1,\ldots,k\}} T_i^{\mathrm{UP}}, \quad \text{with } T_i^{\mathrm{UP}} = (\hat{\pi}_i - \hat{\pi}_0)/\sqrt{\hat{\pi}_i(1-\hat{\pi}_i)/n_i + \hat{\pi}_0(1-\hat{\pi}_0)/n_0}.$$

3. A statistic with pairwise pooled variance estimation

$$T^{\mathrm{PO}} = \max_{i\in\{1,\ldots,k\}} T_i^{\mathrm{PO}}, \quad \text{with } T_i^{\mathrm{PO}} = (\hat{\pi}_i - \hat{\pi}_0) \Bigg/ \sqrt{\hat{\pi}_{0i}(1-\hat{\pi}_{0i}) \Bigg/ \left(\frac{1}{n_i} + \frac{1}{n_0}\right)}$$

and

$$\hat{\pi}_{0i}(r_0 + r_i)/(n_0 + n_i).$$

4. A statistic with total pooled variance estimation

$$T^{\mathrm{GP}} = \max_{i\in\{1,\ldots,k\}} T_i^{\mathrm{GP}}, \quad \text{with } T_i^{\mathrm{GP}} = (\hat{\pi}_i - \hat{\pi}_0) \Bigg/ \sqrt{\hat{\pi}_\bullet(1-\hat{\pi}_\bullet) \Bigg/ \left(\frac{1}{n_i} + \frac{1}{n_0}\right)}$$

and

$$\hat{\pi}_\bullet = r_\bullet/n_\bullet$$

For each of them we investigated the conditional distribution, which fixes the total number of successes $r_\bullet$, and the unconditional distribution, which is the basic assumption for the approximation arguments used for the application of the multivariate normal distribution.

2.1. Conditional distribution of the test statistics

For the exact conditional distribution, $r_\bullet$ is a fixed quantity and the distribution of the test statistics T is obtained by permuting all values of a vector $r=(r_0,\ldots,r_k)$ of success in the set $\Gamma_{r_\bullet}$:

$$\Gamma_{r_\bullet}=\left\{r\colon r_j=0,1,\ldots,\min(n_j,r_\bullet),\sum_{j=0}^{k}r_j=r_\bullet\right\}.$$

The exact conditional distribution of a statistic T is given under the null-hypothesis H_0, i.e. $\pi_0=\pi_1=\cdots=\pi_k=\pi,\pi$ arbitrary

$$P(T(r_0,\ldots,r_k)\geqslant t|R_\bullet=r_\bullet)=\sum_{\left\{\substack{r\in\Gamma_{r_\bullet}\\ T(r)\geqslant t}\right\}}\frac{\prod_{j=0}^{k}\binom{n_j}{r_j}}{\binom{n_\bullet}{r_\bullet}}.$$

The exact conditional distribution under the alternative hypothesis H_A, i.e. $\pi_0<\pi_1$, for some $i\in(1,\ldots,k)$

$$P(T(r_0,\ldots,r_k)\geqslant t|R_\bullet=r_\bullet)=\sum_{\left\{\substack{r\in\Gamma_{r_\bullet}\\ T(r)\geqslant t}\right\}}\frac{\prod_{j=0}^{k}\binom{n_j}{r_j}\pi_j^{r_j}(1-\pi_j)^{n_j-r_j}}{\sum_{\{x\in\Gamma_{r_\bullet}\}}\prod_{j=0}^{k}\binom{n_j}{x_j}\pi_j^{x_j}(1-\pi_j)^{n_j-x_j}}.$$

The conditional distribution of all statistics can be calculated by using an algorithm, which produces all $2x(k+1)$ tables with a fixed number of successes $r_\bullet$. Williams (1988) described the algorithm and investigated also the conditional distribution of T^{GP}. In the case of equal sample sizes for the treatment groups, i.e. $n_1=\cdots=n_k$, the distribution of T^{D} is equal to this distribution, since for a fixed total number of successes $r_\bullet$ and if $n_1=\cdots=n_k$ the standardization is a constant and the statistics are strongly increasing in the difference of the success rates.

The statistics T^{UP} and T^{PO} are defined with factors depending on the number of successes per group. Taking now into account that the distribution is only conditioned by the total number of successes $r_\bullet$, then for $k>1$ the statistics T^{UP} and T^{PO} may have different observed values for a fixed difference in success rates T^{D} which is the result from different success rates per group. For example for $k=2$, $n_i=10$ $(i=0,1,2)$, $r_\bullet=12$ and $T^{D}=0.6$, the statistics T^{UP} and T^{PO} may take four different values. The basic property of T^{UP} and T^{PO} is that both depend not only on the maximum success rate difference T^{D}, but also on the pairwise variance. Therefore, they exhaust the α level better, than the statistics T^{GP} and T^{D}.

It has been already demonstrated by Davis (1986), that in the case $k=1$ all statistics possess the same exact conditional distribution. In general, for equal sample sizes

$n = n_0 = n_i$ the statistics T^{UP} and T^{PO} possess the same distribution. Suissa and Shuster (1985) proved for $k = 1$ that the T^{UP} and T^{PO} are monotonic increasing functions of each other, a result that easily gives the distribution of one from the other, more generally. The amount of computation involved in evaluating a conditional distribution is considerable. For further details see Koch (1996).

2.2. Unconditional distribution of the test statistics

Consider now the case of the exact unconditional distribution of a statistic T for an arbitrary vector π_i. The tail area of the distribution under $\mathrm{H_A}$ is the sum over a product of $(k+1)$ binomial terms

$$P(T(r_0,\ldots,r_k)\geqslant t) = \sum_{\left\{\substack{r\in\Gamma:\\ T(r)\geqslant t}\right\}} \prod_{j=0}^{k} \binom{n_j}{r_j} \pi_j^{r_j}(1-\pi_j)^{n_j-r_j} \tag{2.1}$$

with $\Gamma = \bigcup_{r=0}^{n_\bullet} \Gamma_r$ and where the sum is restricted by values t of the statistic. Under $\mathrm{H_0}$ the tail area of the distribution is calculated by replacing the π_i's with a common π in the above expression. Eq. (2.1) makes clear, that the unconditional tail area depends under $\mathrm{H_0}$ on an unknown parameter π. Therefore, it is necessary to evaluate the tail area, $P(T(R_0=r_0,\ldots,R_k=r_k)\geqslant t|\pi)$, for a fixed value of π. The supremum is taken over π to eliminate the nuisance parameter π. A single α-level critical value of the unconditional test is defined as $t_\alpha = \sup_{\pi\in[0,1]}\{t_\alpha(\pi)\}$, where $t_\alpha(\pi) = \min\{t|P_{\mathrm{H_0}}(T\geqslant t|\pi)\leqslant\alpha\}$ is the critical value for a specific π.

To simplify the further calculation, we assume T to be the maximum of pairwise comparisons S between the control group and k treatment groups, i.e. $T(R_0=r_0,\ldots,R_k=r_k) = \max_{i=1,\ldots,k} S(R_i,R_0)$. Then the tail area of the null distribution of T is given in the Lemma 2.1.

Lemma 2.1. *If S is a function from $\Re^2$ into $\Re^1$ and $(R_0,\ldots,R_k)$ is a vector of independent binomial distributed random variables with $R_i \sim B(n_i,\pi_i)$, then* $\max_{j=1,\ldots,k}(S(R_j,R_0))$ *has the distribution*

$$P\left(\max_{j=1,\ldots,k}(S(R_j,R_0)>t)\right) = 1-\left(\sum_{r_0=0}^{n_0} P(R_0=r_0)\left(\prod_{j=1}^{k}\left(\sum_{S(r_j,r_0)\leqslant t} P(R_j=r_j|R_0=r_0)\right)\right)\right).$$

Proof. Consider the tail area $C = \{(r_0,\ldots,r_k)\in\Gamma:\ \max_{j=1,\ldots,k}(S(R_j,R_0)>t)\}$. Since we assume a pairwise comparison, we have

$$C = \Gamma - \lfloor\{(r_0,\ldots,r_k)\in\Gamma:\ S(R_1,R_0)\leqslant t\}\cap\cdots\cap\{(r_0,\ldots,r_k)\in\Gamma: S(R_k,R_0)\leqslant t\}\rfloor$$

together with the conditioning to the random variable R_0 the result follows directly.

Under the null-hypothesis and for equal sample sizes for each group the distribution becomes

$$P\left(\max_{j=1,\ldots,k}(S(R_j,R_0)>t)|\mathrm{H}_0\colon \pi=\pi_j,\ j=0,\ldots,k\right)$$
$$=1-\left(\sum_{r_0=0}^{n}\left(\binom{n}{r_0}\pi^{r_0}(1-\pi)^{n-r_0}\left(\sum_{S(r,r_0)\leqslant t}\binom{n}{r}\pi^{r}(1-\pi)^{n-r}\right)^{k}\right)\right). \qquad \square$$

This result will be used in Section 3 to calculate exact probabilities and critical constants for the statistics T^{UP} and T^{PO}. It is necessary to give for each function S individual summation limits, which depend on r_0.

Besides the above-described approach to eliminate the nuisance parameter π, we generalized also the methodology described by Suissa and Shuster (1985), which aims to eliminate the effect of the nuisance parameter by maximizing the size of the test over the domain of the nuisance parameter (Basu, 1977). A more straightforward method has been used by Haber (1987). We follow Suissa and Shuster's method and define the function:

$$g(\pi)=P\left(\max_{j=1,\ldots,k}(S(R_j,R_0)\geqslant t_0)|\mathrm{H}_0\colon(\pi,\ldots,\pi)\right)$$
$$=1-\sum_{r_0=0}^{n_0}b(n_0,\pi,r_0)\left[\prod_{j=1}^{k}\left(\sum_{\substack{(r_j r_0)\\ S(r_j r_0)\leqslant t_0}}b(n_j,\pi,r_j)\right)\right] \qquad (2.2)$$

as a tail area probability for a fixed value t_0 under null-hypothesis H_0, where $b(n_j,\pi,r)$ denotes the binomial probability function. The method by which $g(\pi)$ is maximized uses the calculation of the derivate and gives for any interval $I=(a,b)$ an absolute bound M for $|g'(\pi)|$. This result is of Lemma 2.2.

Lemma 2.2. *Let S be a function from $\mathfrak{R}^2$ into $\mathfrak{R}^1$ and $(R_0,\ldots,R_k)$ a vector of independent binomially distributed random variables with $R_i\sim B(n_i,\pi_i)$. The function $g(\pi)$ as defined in Eq. (2.2) has a finite derivative $g'(\pi)$ and an absolute bound M with $|g'(\pi)|<M$, for each $\pi\in[a,b]\subset[0,1]$, $[a,b]$ arbitrary.*

Proof (*See Appendix A*). Finally, let $I_1=[0,0.01]$, $I_2=[0.01,0.02],\ldots,I_{50}=[0.49,0.50]$. For each $I_j=1(1)50$, we can now use Lemma 2.2 to find a bound M_j such that $|g'(\theta_j)|<M_j$ for all $\theta_j\in I_j$ and we can conclude by the mean value theorem of calculus, that

$$g(\theta_j)\in\lfloor g(\pi_j)-0.005M_j,g(\pi_j)+0.005M_j\rfloor,$$
$$\text{for all } \theta_j\in I_j,\ \text{where } \pi_j=(j-0.5)/100.$$

The function $g(\pi)$ can be bounded above by $g(\pi)<\max_{j=1,\ldots,50}\{g(\pi_j)+0.005M_j\}$. Suissa and Shuster's (1985) described a method which gave bounds of specified precision δ. $\square$

3. Computation of exact *p*-values and critical constants for unconditional distributions

3.1. P-value for the exact unconditional distribution

Let S be defined as in Section 2.2, k, n_0 and $n=n_i$, for $i=1,\ldots,k$ be given parameters and t_0 a value of $\max_{j=1,\ldots,k}(S(R_j,R_0))$ for a realization of $(R_0,\ldots,R_k)$. For each $i=0,\ldots,n_0$ and t_0 compute a limit index $l(i,t_0)$ with $S(j,i)<t_0$ for each $j=0,\ldots,l(i,t_0)$. Afterwards define L as a minimum index with $l(j,t_0)=n_1$ for each $j\geqslant L$. Then Eq. (2.2) becomes

$$P\left(\max_{i=1,\ldots,k} S(R_j,R_0)\geqslant t_0|\pi\right)=B(n_0,\pi,L-1)-\sum_{i_0=0}^{L} b(n_0,\pi,i_0)\left[\sum_{i_1=0}^{l(i_0,t_0)} b(n_1,\pi,i_1)\right]^k. \tag{2.3}$$

In the next step evaluate for each $\pi=0.00(0.01)0.99$ the above term $P(\cdot|\pi)$. Observe that the first term $B(\cdot)$ is the cumulative binomial distribution in n_0 and the second term is a weighted sum of products of cumulative binomial distribution in n_1. Finally, the p-value is equal to $p=\max_\pi(P(\cdot|\pi))$.

3.2. Critical constants for unconditional distributions

The above procedure was also used to calculate the critical constants for the statistics T^{UP} and T^{PO}. A critical constant $t^{\mathrm{UP}}_{\max,k,1-\alpha}$ or $t^{\mathrm{PO}}_{\max,k,1-\alpha}$, respectively, for a given k and a level-α test, satisfies the equation

$$t^{S}_{\max,k,1-\alpha}=\inf\left\{t\,\middle|\,\sup_{\pi}\left(P\left(\max_{i=1,\ldots,k} S(R_i,R_0)\geqslant t|\pi\right)\leqslant\alpha\right)\right\}.$$

The starting value for the computation was the $(1-\alpha)$ quantile of the standard normal distribution. Increments were chosen in steps of 0.01 until an attained size satisfied $\alpha^*<\alpha$. Critical constants for T^{PO} and T^{UP} are given in Table 1 for $n=10(1)20(10)100$, $k=2$–5 and $\alpha=0.05$. The attained size α^* is given only once, since for equal sample sizes the distributions of both statistics are the same. For $k=1$ this was shown by Suissa and Shuster (1985), and they gave a detailed table with critical constants for various α and sample sizes. One observation from Table 1 is that as sample size increases the critical values do not decrease monotonically, due to the discrete nature of the binomial distribution. Table 2 gives some examples for different sample sizes. The difference between the exact critical constants for T^{PO} and T^{UP} are pronounced.

Fortran 77 algorithms for quantiles and p-values can be obtained via e-mail (hothorn@ifgb.unihannover.de).

Table 1
One-sided critical values and attained level α^* for T^{UP} and T^{PO} for the exact unconditional distribution with equal sample sizes

n	$k=2$			$k=3$			$k=4$			$k=5$		
	α^*	$t^{PO}_{0.95}$	$t^{UP}_{0.95}$	α^*	$t^{PO}_{0.95}$	$t^{UP}_{0.95}$	α^*	$t^{PO}_{0.95}$	$t^{UP}_{0.95}$	α^*	$t^{PO}_{0.95}$	$t^{UP}_{0.95}$
10	0.0385	1.96	2.17	0.0307	2.25	2.60	0.0393	2.25	2.60	0.0472	2.25	2.60
11	0.0478	1.92	2.10	0.0477	2.17	2.45	0.0444	2.22	2.51	0.0351	2.29	2.63
12	0.0415	2.05	2.26	0.0386	2.20	2.45	0.0490	2.20	2.45	0.0454	2.25	2.53
13	0.0448	1.99	2.16	0.0421	2.18	2.41	0.0477	2.22	2.46	0.0403	2.37	2.67
14	0.0479	1.94	2.08	0.0465	2.17	2.37	0.0465	2.27	2.51	0.0445	2.30	2.55
15	0.0404	2.00	2.14	0.0398	2.20	2.40	0.0404	2.24	2.45	0.0485	2.24	2.45
16	0.0456	1.97	2.10	0.0433	2.14	2.31	0.0435	2.20	2.38	0.0406	2.29	2.51
17	0.0430	2.07	2.21	0.0473	2.13	2.28	0.0464	2.16	2.32	0.0482	2.26	2.45
18	0.0442	2.02	2.14	0.0387	2.13	2.27	0.0491	2.13	2.27	0.0444	2.34	2.54
19	0.0466	1.98	2.09	0.0489	2.10	2.23	0.0464	2.18	2.45	0.0481	2.29	2.46
20	0.0490	1.94	2.04	0.0496	2.09	2.21	0.0494	2.22	2.37	0.0468	2.29	2.46
30	0.0471	2.02	2.09	0.0462	2.11	2.17	0.0465	2.20	2.29	0.0463	2.33	2.44
40	0.0460	2.02	2.07	0.0493	2.10	2.13	0.0495	2.24	2.31	0.0500	2.27	2.34
50	0.0447	2.01	2.05	0.0480	2.09	2.13	0.0483	2.22	2.27	0.0495	2.26	2.32
100	0.0486	1.98	2.00	0.0476	2.13	2.15	0.0495	2.18	2.21	0.0496	2.28	2.30

Table 2
One-sided critical values and attained level α^* for T^{UP}_{max} and T^{PO}_{max} for the exact unconditional distribution with unequal sample sizes

n_0	n_i	$k=2$				$k=3$				$k=4$			
		α^*	$t^{PO}_{0.95}$	α^*	$t^{UP}_{0.95}$	α^*	$t^{PO}_{0.95}$	α^*	$t^{UP}_{0.95}$	α^*	$t^{PO}_{0.95}$	α^*	$t^{UP}_{0.95}$
10	5	0.0403	2.15	0.0272	3.17	0.0263	2.24	0.0396	3.17	0.0337	2.24	0.0490	3.76
20	10	0.0436	2.08	0.0464	2.59	0.0453	2.20	0.0378	2.93	0.0410	2.43	0.0491	2.93
30	15	0.0423	2.05	0.0334	2.74	0.0449	2.33	0.0485	2.74	0.0450	2.35	0.0335	3.03
40	20	0.0471	2.04	0.0330	2.66	0.0480	2.28	0.0481	2.66	0.0446	2.32	0.0383	2.92
50	25	0.0461	2.03	0.0350	2.62	0.0499	2.25	0.0486	2.78	0.0498	2.30	0.0413	2.86
100	50	0.0494	2.02	0.0393	2.53	0.0391	2.26	0.0378	2.75	0.0500	2.26	0.0482	2.75

4. Comparisons with other procedures

In this section we compare the procedure for the unconditional distribution with known alternatives.

A well-known solution to our model setting is the application of the multivariate t-distribution (see Bristol, 1993; Chuang-Stein and Tong, 1995). This application assumes the unconditional probability model. Bristol (1993) considered T^{PO}, T^{UP} and T^{GP}, the latter one was investigated also by Chuang-Stein and Tong (1995). It is a common practice to use Dunnett's critical values, which are denoted by $d_{k,df,\rho,\alpha}$ in this paper and which are based on the k-dimensional multivariate t-distribution. It is not the intention of this paper to argue whether this practice is theoretically valid. Apart

from this, a direct comparison of Dunnett's critical values with our exact critical values may characterize the difference between both sets of values. Consider for $k = 2$ and $\alpha = 0.05$ the equal group sample sizes of 10, 20, 30, 40, 50 and 100. Then the Dunnett's critical values are 2.00, 1.95, 1.94, 1.93, 1.93, 1.92; when $k = 5$, for the same group sizes the values are 2.22, 2.19, 2.18, 2.18, 2.17, 2.17 (Hochberg and Tamhane, 1987). A comparison with the exact critical values given in Table 1 shows for both $k = 2$ and 5 that the exact critical constants for T^{PO} and T^{UP} are generally higher than the Dunnett's critical values. Although the differences are small, it is of interest to know the actual significance level, when Dunnett's critical values are used. An evaluation of this question was done by simulations. For simulations we generated 10 000 samples for $\pi = 0.00(0.01)0.99$ and a configuration discussed by Bristol (1993), which is $k = 2$ and $n_0 = 40$, $n_i = 10$, respectively. The result is shown in Fig. 1. For all three statistics we have used the critical value $d_{2.60,0.45,0.05} = 1.98$ and the exact critical values.

Obviously, the applications of Dunnett's critical values mislead for all three statistics to liberal tests, especially for unequal group sample sizes. If instead exact critical values are used $t^{PO}_{2,40,10,0.95} = 2.09$ and $t^{UP}_{2,40,10,0.95} = 3.66$, then the exact approaches lead to inferences where actual significance level is less than the nominal level. The result is also shown in Fig. 1. The graph for T^{GP} is not shown, since our methodology is not applicable to this statistic. The simulations lead to the conclusion that for unequal group sample sizes the replacement of the exact critical constants by the critical values of the multivariate t-distribution leads to liberal tests.

Another approach by Chuang-Stein and Tong (1995) estimated critical values using simulations for T^{GP} under the assumption of equal sample sizes and a single fixed π. This methodology has only a limited scope of application due to the strict assumptions.

Instead of finding the max-value of the critical constant over π we could use an estimate of π. According to the work of Storer and Kim (1990) for the 2×2 tables the unknown π is estimated from the data by $\hat{\pi}_\cdot = r_\cdot / n_\cdot$. The unconditional distribution with an estimated success rate is

$$P(T(r_0, r_1, \ldots, r_k) \geqslant t | \hat{\pi}) = \sum_{\substack{T(r_0, r_1, \ldots, r_k): \\ T(r_0, r_1, \ldots, r_k) \geqslant t}} b(r_0, n_0, \hat{\pi}) \cdot b(r_1, n_1, \hat{\pi}) \cdot \ldots \cdot b(r_k, n_k, \hat{\pi}). \tag{2.4}$$

The null-hypothesis is rejected if $P(T(r_0, r_1, \ldots, r_k) \geqslant t | \hat{\pi}) \leqslant \alpha$.

Westfall and Young (1989) described resampling-based tests for the comparison of multivariate binomial observations from k groups. They discussed the bootstrap and permutation resampling and applied both approaches with different analysis techniques (e.g. Fisher's exact test, standardized difference of frequencies). Beside univariate also multiple testings procedures have been developed, for details we refer to the textbook of Westfall and Young (1993). These methods are currently available in the SAS/STAT MULTTEST procedure. In Chapter 6 we give a numerical example and analyze the data also using the SAS procedure MULTTEST.

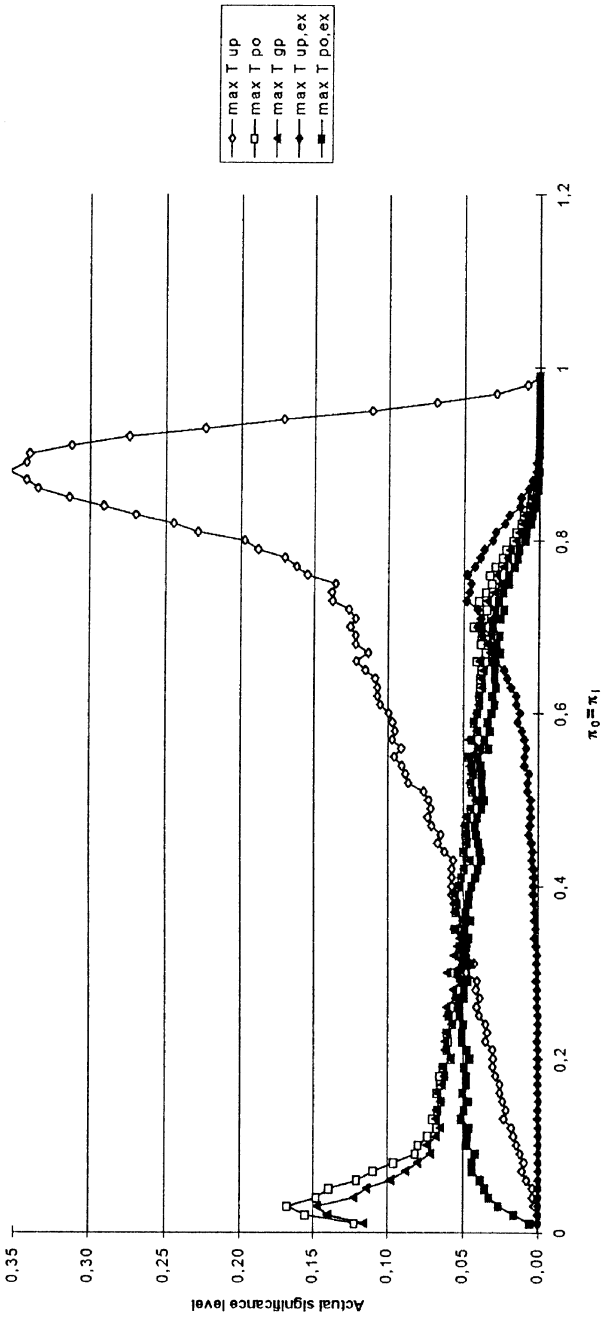

Fig. 1. Actual significance levels of one-sided 0.05 nominal level tests using multivariate t-distribution and the exact unconditional distribution for $k = 2$, $n_0 = 40$, $n_i = 10$.

5. Application in a stepwise multiple comparison procedure

The exact unconditional distribution of T^{PO}_{max} and T^{UP}_{max} can be used in a multiple step-down test procedure. We consider the problem of testing $k \geqslant 2$ hypotheses, $H_1, \ldots, H_k$, to control the type I FWE. The simultaneous multiple step down procedure (SD) for binary data operates as in the case of the normal distribution (Dunnett and Tamhane, 1991). Let $c_m = t_{m,n_0,n_1,\ldots,n_m,\alpha}$ be the upper α critical constant of $T^{max}_m = \max_{i=1,\ldots,m} S_i$ for $m = 1, \ldots, k$ and order the statistics $t_{(1)} \leqslant \cdots \leqslant t_{(k)}$. Denote the corresponding hypotheses with $H_{(1)}, \ldots, H_{(k)}$, and then reject any $H_{(i)}$ if $H_{(j)}$ is rejected for $j = k, \ldots, i$, i.e. $t_{(j)} \geqslant c_j$ for $j = k, \ldots, i$. The closed test procedure can be applied to the system of T^{max}_m and it becomes clear that the multiple step down procedure controls the type I FWE.

As an alternative method Piegorsch (1991,1995) suggested to apply an α-adjustment procedure to two sample tests. In principle each two sample test, like Fisher's exact test and Barnard's test, could be used together with an α-adjustment procedure, for example Bonferroni, Bonferroni–Holm (Holm, 1977) or Hommel (1988).

A comparison of the simultaneous multiple step down procedure and the Bonferroni–Holm α-adjustment procedure is given in Table 3 for equal sample sizes. The table gives the exact unconditional multivariate critical values $t^{PO}_{k,\alpha=0.05,ex}$ and the stepwise α-adjusted exact unconditional critical values $t^{PO}_{1,\alpha=0.05/k,ex}$ for a many to-one comparison. A difference in favor of the simultaneous approach occurs when the number of treatment groups $k \geqslant 3$ and the sample sizes are greater than 20 with $t(SD) = t^{PO}_{k,\alpha=0.05,ex}$ and $t(BH) = t^{PO}_{1,\alpha=0.05/k,ex}$.

For both procedures the actual significance level is less than nominal level. To demonstrate this property, we did simulations for different profiles of the success rates. For simulations, we generated 10 000 samples of binomial responses and evaluated the data for any false positive conclusion. Table 4 gives the empirical type I error rates, with maximal standard error of 0.2%.

We notice that in all cases the observed empirical type I error rate is less than 5%. Comparing first the α adjustment procedures Hommel's step-up procedure exhaust the

Table 3
One-sided critical constants t for simultaneous (SD) and α-adjusted step down (BH) procedures using the exact unconditional distribution ($\alpha = 0.05$)

$n_0 = n_i$	Procedure	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$
10	t(SD)	1.79	1.96	2.25	2.35	2.35
	t(BH)	1.79	1.96	2.25	2.35	2.35
20	t(SD)	1.78	1.94	2.09	2.22	2.29
	t(BH)	1.78	1.99	2.22	2.29	2.38
50	t(SD)	1.69	2.01	2.09	2.21	2.26
	t(BH)	1.69	2.01	2.21	2.26	2.40
100	t(SD)	1.70	1.98	2.13	2.18	2.28
	t(BH)	1.70	1.99	2.17	2.28	2.34

Table 4
The empirical overall type I error rate (%) for the α adjustment procedure according to Bonferroni–Holm and Hommel applied to Fisher's exact test and Barnard's test and simultaneous multiple step down procedure (SD)

Success rates					n_0	n_1	Empirical overall type error rate (%)						
π_0	π_1	π_2	π_3	π_4			Bonferroni–Holm			Hommel			SD
							T_{ex}^{FI}	$T_{mid\,p}^{FI}$	T_{ex}^{PO}	T_{ex}^{FI}	$T_{mid\,p}^{FI}$	T_{ex}^{PO}	$T_{ex,k}^{PO}$
0.05	0.05	0.05	–	–	20	20	0.2	0.2	1.1	0.2	0.2	1.2	1.1
					40	40	0.4	0.5	3.6	0.5	0.6	4.2	3.6
					80	80	1.6	3.0	3.8	1.7	3.0	4.1	4.5
0.20	0.20	0.20	–	–	20	20	2.2	4.3	4.7	2.3	4.3	5.0	4.7
					40	40	2.6	3.5	3.6	2.7	3.7	3.8	3.6
					80	80	2.7	3.5	4.0	2.9	3.7	4.1	4.3
0.05	0.05	0.05	0.05	–	20	20	0.0	0.0	0.3	0.0	0.0	0.3	1.7
					40	40	0.6	0.7	2.2	0.6	0.8	2.6	2.2
					80	80	1.7	2.7	4.0	1.9	2.9	4.5	4.0
0.20	0.20	0.20	0.20	–	20	20	1.6	2.5	2.8	1.7	2.8	3.2	3.8
					40	40	1.9	3.5	4.1	2.2	3.8	4.4	4.5
					80	80	2.7	3.1	3.7	2.9	3.5	4.0	4.6
0.05	0.05	0.05	0.05	0.05	20	20	0.1	0.1	0.4	0.1	0.1	0.4	0.4
					40	40	0.2	0.3	2.5	0.2	0.3	3.1	2.5
					80	80	1.2	2.1	2.9	1.4	2.3	3.2	3.0
0.20	0.20	0.20	0.20	0.20	20	20	1.5	3.2	3.5	1.7	3.5	3.9	3.5
					40	40	1.8	3.1	3.4	2.2	3.4	3.6	3.4
					80	80	2.6	3.5	3.8	2.9	4.0	4.2	4.2

T_{ex}^{FI}=Fisher's exact test, $T_{mid\,p}^{FI}$=Fisher's exact test with mid-p value (Hirji et al., 1991), T_{ex}^{PO}=Barnard's test.

nominal level better than the Bonferroni–Holm step down. The largest type I error rates for these procedures are observed for the Barnard's test T_{ex}^{PO}. The simultaneous step-down procedure (SD) $T_{ex,k}^{PO}$ has always error rates greater than the Bonferroni–Holm, but sometimes less than Hommel's step-up with Barnard's test T_{ex}^{PO}.

6. An example

As an illustration consider the result of an experiment in toxicology. The experiment compared three doses (10, 50, 100 mg/kg/day) of a new pharmaceutical compound to negative control C when applied to mice during a long-term administration. Of interest was the survival rate after 6 months (Hothorn, 1994), whereas a many-to-one procedure without order restriction was used due to possible downturn phenomena at high doses. The total number of mice observed was $n_{\cdot} = 100$ and the within group sample sizes were $n_0 = 40$ and $n_i = 20$ for $i = 1, 2, 3$. The results appear in Table 5.

The maximum difference between the negative control C and a treatment group occurs for the highest dose level of 100 mg/kg daily, therefore the maximum statistic is $T_{max}^{PO} = 2.74$, the multivariate exact unconditional distributions gives a p-value of

Table 5
Chronic toxicity study in mice over six months

	C	10 mg/kg	50 mg/kg	100 mg/kg
Total number of mice (n_i)	40	20	20	20
Total number of deaths (r_i)	4	1	6	8
$\hat{\pi}_i$	0.10	0.05	0.30	0.40
T_i^{PO}		−0.66	1.96	2.74

Table 6
Multiple comparison procedures: Bonferroni–Holm (BH) with Barnard's test and simultaneous step-down procedure (SD) based on the exact unconditional binomial distribution

Step	Ordered		Barnard's test		SD exact uncond. Simultaneous p-value
	Comparison	Statistic	Raw p-value	BH p-value	
1	C vs. 100	2.74	0.005	0.016	0.009
2	C vs. 50	1.96	0.040	0.080	0.052
3	C vs. 10	−0.66	0.606	0.606	0.606

Table 7
Permutation test based Fisher's exact and bootstrap test based on t-test with Bonferroni–Holm procedure from SAS PROC MULTTEST

Step	Ordered Comparison	Fisher's exact		Permutation BH p-value	t-test based on T^{GP}		Bootstrap
		Raw p-value	BH p-value		Raw p-value	BH p-value	BH p-value
1	C vs. 100	0.010	0.029	0.013	0.002	0.007	0.010
2	C vs. 50	0.059	0.117	0.068	0.027	0.055	0.058
3	C vs. 10	0.880	0.880	0.789	0.686	0.686	0.680

$p = 0.013$, whereas the p-value for the exact two samples according to Banard's test gives a p-value of $p = 0.005$. In the next step we apply the Bonferroni–Holm procedure and the simultaneous step-up procedure. The following Table 6 gives for each step the observed p-values and the individual α-levels.

According to both procedures only the first comparison between C and 100 mg/kg daily leads to a statistically significant result at the multiple 5%-level, since in the second step for the Bonferroni–Holm procedure the adjusted p-value is 0.080 and for the simultaneous step-down procedure the observed alpha of 0.052 is greater than 0.050, but C vs. 50 can be rejected at the 0.052 level with the simultaneous step-down procedure.

The following Table 7 gives the result of two resampling analysis using the SAS PROC MULTTEST. Firstly, a permutation test is performed, which uses one-sided p-values of Fisher's exact test (raw p-values) and their adjustment according to the Bonferroni–Holm procedure. These latter p-values are used as thresholds in the permutation test. Secondly, a bootstrap test is calculated which is based on the

Bonferroni–Holm adjusted p-values of a t-test which uses the statistic T^{GP}. The permutation test in the Bonferroni–Holm procedure gives in all steps greater p-values than the bootstrap based on the t-test. The latter are still larger than the p-values of the simultaneous step-down procedure in Table 6, but the difference is marginal.

We conclude this section by giving the statements for the analysis of the above example using SAS PROC MULTTEST:

```
DATA num_dat; INPUT c n s; CARDS;
1 40 4 ... ;
DATA trans_dat; SET num_dat; DO I=1 to n; y=i LE s; OUTPUT; END; RUN;
PROC MULTTEST DATA = trans_dat STEPBON STEPPERM; CLASS c;
TEST FISHER(y/UPPERTAILED);
CONTRAST '1 vs 2'–1 1 0 0;
CONTRAST '1 vs 3'–1 0 1 0;
CONTRAST '1 vs 4'–1 0 0 1;
RUN;
```

(The Number of resamples was 20 000 and the seed number 35 409, replace FISHER by MEAN and STEPPERM by STEPBOOT for the bootstrap test based on the t-test).

7. Two-sided tests

All of the above results can be extended to two-sided tests. Then the hypotheses are $H_{0i} : \pi_i = \pi_0$ versus $H_{Ai} : \pi_i \neq \pi_0$, $i \in (1,\ldots,k)$. The statistics from Section 2 becomes $\max_{1\leqslant i\leqslant k}|T_i|$. For the conditional distribution one has to bear in mind that two different definitions for a probability measure are possible (see Hirji et al., 1991). The p-values and critical values for the unconditional distribution can be computed by changing of the summation limits in Corollary 2.1 and Eq. (2.3) to $|T(r_0, r_i)| \leqslant t$.

8. Extensions and concluding remarks

In this article we have focused on the case with equal sample sizes within the k treatment groups. This restriction was used to derive from Eq. (2.2) their simplification, which is given in Eq. (2.3). with the latter one, p-values and critical constants were easily calculated. The general situation with unequal sample sizes is not covered in this article.

Further investigations should focus on a comparison of the acceptance regions for the different statistics, as discussed by Storer and Kim (1990) for $k = 1$, including their test statistic defined in Eq. (2.4). Of course the convergence speed to the exact unconditional distribution would be of primary interest.

In Section 5 we have demonstrated that the application on other multivariate distribution in a step-down procedure hardly competes with the Bonferroni–Holm procedure applied to Barnard's test. For $k = 2$ no real advantage of the multivariate statistic could be detected. The SAS/PROC MULTTEST offers a good possibility to calculate Bonferroni–Holm adjusted p-values, but the procedure has still the disadvantage of using for the permutation test the rawvalue of the Fisher's exact test and in the unconditional model the bootstrap is based on the asymptotic distribution.

Acknowledgements

The authors thank Ajit Tamhane and Yoav Benjamini for their insightful comments. Moreover, the authors thank Jeroen Tolboom for bringing to their attention the evaluation using resampling methods in SAS. Additionally, the editors and referees provided a number of helpful comments leading to an improved final version.

Appendix A

We use the following definitions and equations:

Let $b(\pi, n, i) = \binom{n}{i} \pi^i (1-\pi)^{n-i}$, the derivative with respect to π is:

$b'(\pi, n, i) = nb(\pi, n-1, i-1) - nb(\pi, n-1, i)$ and the set C as defined above. The derivative of $g(\pi)$ is given by the following equation:

$$\begin{aligned} g'(\pi) = \sum_{i_0=0}^{n} & \left[nb(\pi, n-1, i_0) \left[\left(\sum_{(i_0,j)\in C^c} b(\pi, n, j) \right)^k \right] - nb(\pi, n-1, i_0-1) \right. \\ & \left. \times \left[\left(\sum_{(i_0,j)\in C^c} b(\pi, n, j) \right)^k \right] \right] \\ & + \sum_{i_0=0}^{n} \left[\sum_{(i_0,j)\in C^c} nb(\pi, n, i_0) b(\pi, n-1, j) - nb(\pi, n, i_0) b(\pi, n-1, j-1) \right] \\ & \times \left[k \left[\sum_{(i_0,j)\in C^c} b(\pi, n, j) \right]^{k-1} \right]. \end{aligned}$$

For the last equation the following partial sums are defined:

$$\begin{aligned} g_1'(\pi) = \sum_{i_0=0}^{n} & \left[nb(\pi, n-1, i_0) \left[\left(\sum_{(i_0,j)\in C^c} b(\pi, n, j) \right)^k \right] - nb(\pi, n-1, i_0-1) \right. \\ & \left. \times \left[\left(\sum_{(i_0,j)\in C^c} b(\pi, n, j) \right)^k \right] \right] \end{aligned}$$

$$g_2'(\pi) = \sum_{i_0=0}^{n} \left[\sum_{(i_0,j)\in C^c} nb(\pi,n,i_0)b(\pi,n-1,j) - nb(\pi,n,i_0)b(\pi,n-1,j-1) \right] \times \left[k \left[\sum_{(i_0,j)\in C^c} b(\pi,n,j) \right]^{k-1} \right].$$

Consider now the boundary of the C^C for $g_1'(\pi)$, by

$$W = \{(r_1,r_2):(r_1,r_2) \in C^C \text{ and } (r_1+1,r_2) \notin C^C\},$$

only for indices (r_1,r_2) of this set W differences for the two terms exist. After cancellation of identical terms with different signs $g_1'(\pi)$ becomes

$$g_1'(\pi) = \sum_{i_0=0}^{n} \left[nb(\pi,n-1,i_0) \left[\left(\sum_{(i_0,j)\in W} b(\pi,n,j) \right)^k \right] \right] = \sum_{(i_0,j)\in W} nb(\pi,n-1,i_0)b(\pi,n,j)^k.$$

The other boundary is $V = \{(x_1,x_2):(x_1,x_2) \in C \text{ and } (x_1,x_2-1) \notin C\}$.

This leads to

$$g_1'(\pi) = \sum_{(i_0,j)\in V} \left[n \binom{n}{i} \binom{n-1}{j-1} [(1-\pi)^{i+j-1}(\pi)^{2n-i-j}]^k \times \left[\binom{n}{i} (1-\pi)^i(\pi)^{2n-i} \right]^{k-1} \right].$$

A similar argument holds for $g_2'(\pi)$:

$$g_2'(\pi) = \sum_{(i_0,j)\in V} -\binom{n}{i} \binom{n-1}{j-1} \pi^{i+j-1}(1-\pi)^{2n-i-j} k \left[\binom{n}{j} (\pi)^j(1-\pi)^{n-j} \right]^{k-1}.$$

In total the derivative of $g(\pi)$ can finally be written as

$$g'(\pi) = \sum_{V} n \binom{n}{i} \binom{n-1}{j-1} \left\{ \left[k \binom{n}{j}^{k-1} \pi^{i+kj-1}(1-\pi)^{(k+1)n-i-kj} \right] - \left[\binom{n}{j}^{k-1} (\pi)^{(k+1)n-ki-j}(1-\pi)^{ki+j-1} \right] \right\}. \quad \text{(A.1)}$$

For $r = i + kj - 1$, $r = ki + j - 1$ and $s = n(k=1) - 1$ the derivative is a linear combination of terms of the form $h(\pi) = \pi^r(1-\pi)^{s-r}$. Now the original argument of Suissa and Shuster (1985) could be applied. For any given interval $I = (a,b)$ with $0 < a < b < 1$ a supremum and infimum for the function $h(\pi)$ could be easily defined. An upper bound for $g'(\pi)$ is obtained on (a,b) by substituting the suprema in each positive term and the infimum in each negative term of Eq. (A.1). Similarly, a lower bound for $g'(\pi)$ is obtained on (a,b) by reversing the substitution. Finally, a bound M for $|g'(\pi)|$ on (a,b) is taken as the larger of the bounds. Details of the proof are available from the authors.

References

Ahner, C., Passing, H., 1983. Berechnung der multivariaten t-Verteilung and simultane Vergleiche gegen Kontrolle bei ungleichen Gruppenbesetzungen, EDV. Medizin Biologie 14, 113–120.

Basu, D., 1977. On the elimination of nuisance parameters. J. Amer. Statist. Assoc. 72, 355.

Bristol, D.R., 1993. One-sided multiple comparisons of response rates with a control. In: Hoppe, F.M. (Ed.), Multiple Comparisons, Selection and Applications in Biometry. Marcel Dekker, New York, pp. 77–96.

Chuang-Stein, C., Tong, D.M., 1995. Multiple comparisons procedures for comparing several treatments with a control. Statist. Med. 14, 2509–2522.

Davis, L.J., 1986. Exact tests for 2×2 contingency tables. Amer. Statist. 40, 139–141.

Dunnett, C.W., Tamhane, A.C., 1991. A step-up-multiple test procedure. J. Amer. Statist. Assoc. 87, 162–170.

Fairweather, W.R., 1987. Comparing proportion exposed in case-control studies using several control groups. Amer. J. Epidemiol. 126, 170–178.

Haber, M., 1987. A comparison of some conditional and unconditional exact tests for 2×2 contingency tables. Commun. Statist. 16, 999–1013.

Holm, S.A., 1977. A simple sequentially rejective multiple test procedure. Scand. J. Statist. 6, 65–70.

Hommel, G., 1988. A stagewise rejective multiple test procedure based on a modified Bonferroni test. Biometrika 75, 383–386.

Hothorn, G., 1994. Biometrie. In: Marquardt, H. (Ed.), Lehrbuch der Toxikologie. BI Wissenschaftli cher Verlag, Mannheim, pp. 15–31.

Hirji, K.F., Tan, S.-J., Elashoff, R.M., 1991. A quasi-exact test for comparing two binomial proportions. Statist. Med. 10, 1137–1153.

Hochberg, Y., Tamhane, A.C., 1987. Multiple Comparison Procedures. Wiley, New York.

Koch, H.F., 1996. Teststatistiken für die 'many-to-one' Versuchsanlage im Falle dichotomer Ereignisse, Dissertation, Universität Hannover.

Passing, H., 1984. Exact simultaneous comparisons with control in an $r \times c$ contingency table. Biometrical J. 26, 643–654.

Piegorsch, W.W., 1991. Multiple comparisons for analyzing dichotomous response. Biometrics 47, 45–52.

Piegorsch, W.W., 1995. Many-to-one comparison procedures for dichotomous endpoints. In: Vollmar, J. (Ed.), Biometrie in der chemisch-pharmazeutischen Industrie, vol. 6, pp. 70–76.

Storer, B.E., Kim, C., 1990. Exact properties of some exact test statistics for comparing two binomial proportions. J. Amer. Statist. Assoc. 85, 146–155.

Suissa, S., Shuster, J.J., 1985. Exact unconditional sample sizes for the 2×2 binomial trial. J. Roy. Statist. Soc. A 148, 317–327.

Westfall, P.H., Young, S.S., 1989. P-value adjustments for multiple tests in multivariate binomial models. J. Amer. Statist. Assoc. 84, 780–786.

Westfall, P.H., Young, S.S., 1993. Resampling Based Multiple Testing – Examples and Methods for p-value Adjustment. Wiley, New York.

Williams, D.A., 1988. Tests for differences between several small proportions. Appl. Statist. 37, 421–434.