

MODIFICATIONS OF THE CLOSURE PRINCIPLE FOR ANALYZING TOXICOLOGICAL STUDIES

LUDWIG A. HOTHORN, PHD

University of Hannover, Hannover, Germany

A Drug Information Association Workshop on "Statistical Methodology on Non-Clinical and Toxicological Studies" was held on March, 25-27, 1996. The purpose of this meeting was to discuss the appropriateness of current and new biostatistical methods in this field of drug development. This paper proposes a simple closure test for dose-response relationships under real data conditions. This approach takes into account deviation from variance homogeneity and monotonicity assumptions. Moreover, this approach can be easily adapted for "any" kind of two-sample tests, for example, nonparametric, dichotomous, censored, and so forth. Power was compared by a simulation study for selected conditions.

Key Words: MED; Closure principle; Trend test; Toxicological studies

INTRODUCTION

THIS ARTICLE DESCRIBES a simple closure testing procedure for dose-response relationships under real data conditions. Although this approach can be used for randomized clinical trials, pharmacological studies, and several toxicological studies, for example, reproductive toxicity, repeated toxicity studies (1) are considered here for analyzing the dose-response relationship. A one-way design: $\{C-, D_1, \dots, D_k\}$ (with $C-$. . . negative control; D . . . dose group; with k not too small $\in \{2,3,4\}$) is assumed and multiple endpoints of a different scale, for example, approximate gaussian distributed (eg, body weight), a highly skewed distribution (eg, ASAT enzyme), dichotomous (eg, number of animals that died/number of animals

at risk), or ordered categorical (eg, scored histopathological findings). Independent univariate analyses of the dose-response relationship should be performed, however, based on the same methodology. Only model-free approaches should be discussed here, because a priori in such "screening" studies a model for the profile cannot be assumed, even for a single endpoint.

The purpose of toxicological studies is safety assessment by comparing the doses to the concurrent negative control (many-to-one comparisons). Traditionally, the classical hypotheses are used: null-hypothesis for no difference and alternative hypothesis for a difference (or one-sided for an increase). In this setting, type I error α represents producer's risk and type II error β consumer's risk. A direct control of producer's risk (via a priori definition of α) is possible, but control of consumer's risk is of primary concern in risk assessment. Therefore, Hauschke (2,3) proposed the reversing of the hypotheses in the sense of a k -sample equivalence problem. For this approach a threshold δ must be de-

Presented at the DIA Workshop "Statistical Methodology in Non-Clinical and Toxicological Studies," March, 25-27, 1996, Bruges, Belgium.

Reprint address: Prof. Dr. L. A. Hothorn, Herrenhauser Str. 2, D-30419 Hannover, Germany.

fined a priori or a posteriori. In toxicological studies with a primary endpoint, for example, number of revertants, such an approach can be used, as demonstrated for the Ames mutagenicity assay by Neuhäuser and Hothorn (4,5). In the multiple endpoint "screening" studies for repeated toxicity, however, this equivalence approach seems to be difficult. Therefore, the traditional hypotheses will be used here. In a k -sample design, the problem of multiple comparison exists. A general disadvantage of multiple comparison procedures is the loss of power (=increase of type II error) with an increasing number of comparisons. This increase of consumer's risk is exactly what should be avoided in safety studies. Therefore, this paper describes approaches with power as high as possible. Modifications of the closure principle will be used. But this represents a testing approach: a reject/accept decision only. Alternatively, the confidence interval could be used, including more information. This will be not discussed here. Power depends not only on the type of trend test, however, but also on the robustness against violation of the a priori assumptions:

- Variance heterogeneity,
- Lack of balance,
- Unknown shape of the dose-response relationship,
- Possible downturns at high doses, and
- Different endpoints.

For these conditions robust and simple modifications will be presented here.

A global trend test, for example, the well-known Jonckheere (6) trend test, is not sufficient for repeated toxicity studies, because the only decision possible is trend/no trend. Local answers, for example, no observed effect dose (NOED), are also needed. The NOED significance approach used here should be assumed roughly as minimal effective dose (MED)-1. The closure principle according to Marcus et al. (7) is one possibility for combining global and local decisions without additional α effort.

THE CLOSURE PRINCIPLE UNDER ORDER RESTRICTION

Principle

The closure principle according to Marcus et al. (7) contains the following elements:

- Definition of the elementary hypotheses, for example, all-pair comparisons, many-to-one comparisons, comparisons between successive dose steps, and so forth,
- Constructing the closure hypotheses system by partial hypotheses up to the global hypothesis,
- Application of suitable level α test for each hypothesis (the choice is free), and
- Conditionally testing from top to bottom, each at level α .

Definition of the Elementary Hypotheses

Figures 1 and 2 show the closure system for $k = 4$ groups. The comparison of successive dose steps to estimate the highest effective dose step is not relevant in toxicological studies. Closure tests which are relevant for clinical trials were described by Budde and Bauer (8), Bauer and Budde (9), and Rom et al. (10).

Choice of Contrast Tests

Trend tests can be distinguished according to their consideration of the dose levels into the test statistics: direct (respective as a transformed score) or qualitative only. It could be assumed that only qualitative consideration would yield an information loss and consequently a loss in power. This is not, however, true in general. On one hand, the dose levels may not be optimal and better scores should be used (11), and on the other hand dose levels in toxicological studies are often selected in expectation of a linear dose response relationship to either linear or logarithmic dose scale. For the a priori unknown profile of the dose-response relationship, model-based trend tests show a poor power if the preselected model and the data differ.

Moreover, in the parametric context, trend tests can be seen as contrast tests. Barlow et

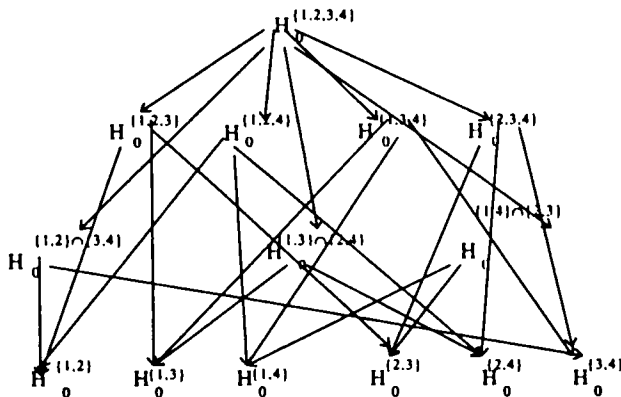


FIGURE 1. All-pair comparisons.

al. (12) demonstrated that the likelihood ratio test is a contrast test with infinity contrasts. Robertson et al. (13) revealed that Williams' (14,15) MLE-test is a special form of multiple contrast test (consisting of a pairwise and several reverse Helmert contrasts). Mehta et al. (16) revealed that the regression approach according to Tukey et al. (11) (a corrected version was published by Capizzi et al. [18]) is also equivalent to multiple contrast tests.

One of the problems in application of trend tests is their reliance on the a priori unknown shape of the dose-response profile. For instance, using the concept of Pitman asymptotic relative efficacy (Ratio of sample sizes to guarantee the same power for $\alpha \Rightarrow 0, \delta \Rightarrow 0, n_i \Rightarrow \infty$, or ratio of the power function in the point α) the efficacy for several contrasts and selected dose-response profile

contrasts in relation to the best power contrast can be compared according to Neu-häuser (17) (Table 1).

From this table it can be concluded that the loss in power for the nonoptimal contrast test depending on the a priori unknown profile is large. Moreover, this loss in power occurs over a broad region of the noncentrality parameter δ . The power of Helmert and reverse Helmert contrasts (19) can be simply computed by noncentral t-distribution. Figure 3 shows for two extreme shapes: convex and concave, the power of both contrasts for $k = 4, n_i = \text{const} = 10$, and $\alpha = 0.05$. It is easy to see that for convex profiles, the Helmert contrast is much more powerful than reverse contrast, and for concave profiles the opposite is true. This power difference is dramatic even for small shifts. It should be made clear that, in principle, the shape of the dose-response relationship is unknown a priori. Several concepts are now used to deal with this problem:

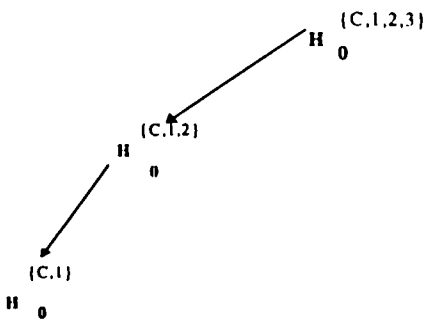


FIGURE 2. Many-to-one comparisons (one-sided ordered alternative hypothesis for MED-estimation).

- Ignoring this important fact: this is a commonly used practice but it is not acceptable in safety studies (due to a dramatic power loss),
- Looking at the data, characterizing a profile type, selection of the best suitable trend test represents one kind of hunting for significance,
- Bartholomew's (20) LRT test: a good version but limited to some ideal conditions,

TABLE 1
Efficacy of Several Contrasts (k = 3)

Contrast	Linear Profile	Convex Profile	Concave Profile
Helmert	0.60	optimal	0.11
Reverse Helmert	0.60	0.11	optimal
Maximin	0.97	0.65	0.65
Linear	optimal	0.60	0.60
Jonckheere-Analagon*	0.84	0.93	0.28
Jonckheere-Reverse-Analagon	0.84	0.28	0.93
Pairwise	0.90	0.67	0.67

*A new form of a parametric contrast according to Neuhäuser (17).

- The multiple contrast test according to Mukerjee et al. (21,22), and
- The adjustive approach according to Neuhäuser (17).

Contrast tests are simply defined by the linear combination of the mean values and univariate distributed:

$$t_C^{\text{parametric}} = \frac{\sum_{i=0}^k c_i \bar{x}_i}{\sqrt{MQ_R \sum_{i=0}^k \frac{c_i^2}{n_i}}} \propto t_{df = \sum_{i=0}^k n_i - (k+1), 1-\alpha}$$

with $\sum_{i=0}^k c_i = 0$.

Researchers have a choice of contrast coefficients; several versions were described in the literature (Table 2). A comparison of some contrast tests using several step-wise proce-

dures was recently published by Tamhane et al. (26). The correlation between two contrasts according to Bechhofer and Dunnett (27) in the variance homogeneity case is:

$$\rho_{a,b} = \frac{\sum_{i=0}^k a_i b_i / n_i}{\sqrt{\sum_{i=0}^k a_i^2 / n_i \sum_{i=0}^k b_i^2 / n_i}}$$

If $\sum_{i=0}^k a_i b_i = 0$ the contrasts are orthogonal and uncorrelated. Numerous publications exist on power comparison of several contrast tests (also with comparison to the other trend tests). The majority of the conclusions about the advantage of a selected contrast test are wrong due to ignoring the a priori unknown shape. It is simply true that if the contrast fits the empirical shape then the test power is optimal, for example, for a linear shape

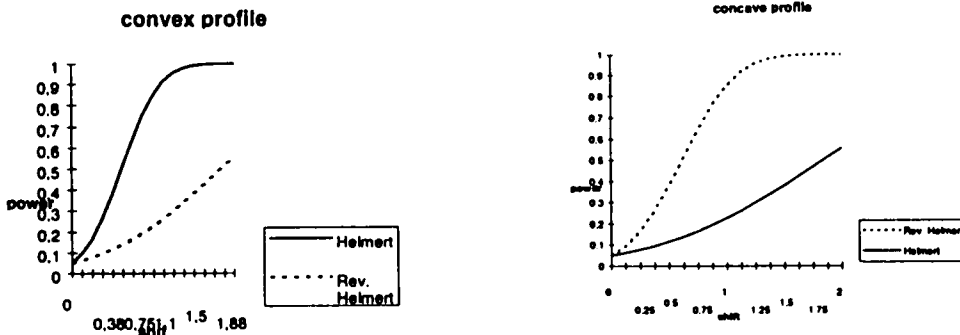


FIGURE 3. Comparison of power functions.

TABLE 2
Selected Contrast Tests

Contrast Type	Reference	Coefficients c_i
Helmert	Ruberg (23)	$c_i = -1, i = 0, 1, \dots, k - 1$ $c_k = k$
Reverse Helmert	Fligner and Wolfe (19)	$c_0 = k$ $c_i = 1, i = 1, 2, \dots, k$
Maximin	Abelson and Tukey (24)	$c_i = \sqrt{i} \left(-\frac{i}{k+1} - \sqrt{(i+1) \left(1 - \frac{i+1}{k+1} \right)} \right)$
Pairwise	Hothorn and Lehmacher (25)	$c_0 = -1$ $c_i = 0, i = 1, 2, \dots, k - 1$ $c_k = 1$
Linear	Rom et al. (10)	$c_0 = -k$ $c_i = c_{i+1} + 2, i = 1, 2, \dots, k$
Jonckheere-Analogue	Neuhäuser (17)	$c_0 = -1$ $c_i = \left(\sum_{j=i+1}^k 1/j \right) - 1, i = 1, 2, \dots, k - 1$ $c_k = \sum_{j=1}^k 1/j$

the linear contrast, for a convex shape the Helmert contrast is optimal, and so forth.

For the sake of simplicity and to allow for generalization to any kind of two-sample pairwise contrasts {C-vs. D_i} will be used here according to Hothorn and Lehmacher (25). This represents a special case of the principle of the a priori ordered test due to its importance according to Maurer et al. (28). The question arises: using such a simple contrast does a power loss exist (ignoring the information between doses)? In the parametric case (k = 4, n = 10) the power can be simply computed using noncentral t-distribution (Table 3). It can be concluded that the pairwise contrast is a simple and average powerful competitor to Helmert or reverse Helmert contrasts (representative of a broad class of alternatives between convex and concave profiles) in the case of an unknown dose-response relationship.

Type of Order Restriction

On the one hand the dose response relationship can be assumed to be strictly monotonic and tests based on a total order restriction

can be used, for example, pairwise contrasts, Helmert contrasts, reverse Helmert contrasts, or the ML-procedure according to Williams (14,15). On the other hand, in toxicology, downturns at high doses are possible. Procedures should be robust against this effect. In this situation MCPs without order restriction can be used, for example, in the parametric case Dunnett's (29) procedure (or its step-wise modifications). Alternatively, partial order restriction can be used, for example, Shaffer's (30) procedure. Based on ordered

TABLE 3
Power of Three Contrast Tests

Expected values	Reverse		
	Helmert	Helmert	Pairwise
0, 0.3, 0.6, 0.9	0.49	0.49	0.63
0, 0.6, 1.2, 1.8	0.94	0.94	0.99
0, 0, 1.8, 1.8	0.94	0.94	0.99
0, 0.9, 0.9, 0.9	0.20	0.78	0.63
0, 1.8, 1.8, 1.8	0.49	0.99	0.99
0, 0, 0, 0.45	0.34	0.11	0.26
0, 0, 0, 0.9	0.78	0.20	0.63
0, 0, 0, 1.35	0.90	0.34	0.79
0, 0, 0, 1.8	0.99	0.49	0.99

p-values $p_{(1)} < \dots < p_{(k)}$ the following decision is made: if the smallest $p_{(i)} < \alpha/k$ reject H_0^i and all hypotheses implied by it, namely: $i + 1, \dots, k$ (no decision is made about these doses). Go on to test the next dose $i - 1$ versus control at level $\alpha/(i - 1)$, and so forth. Rom et al. (10) published an analogous test for linear orthogonal contrasts based on the decision: $\min p_i(\forall \text{contrast tests}) < t_{k,df,\rho,\alpha}$. This approach, however, is numerically difficult and limited in balance and gaussian distribution. For continuous endpoints variance heterogeneity (combined with lack of balance) reduces power much more than typical violations of gaussian distribution for standard procedures based on MQ_R , for example, Dunnett (29) MCP or nonparametric MCP (31) according to Ortseifen and Hothorn (32). Therefore, tests should either be adjusted (eg, Welch-t-test) or the heterogeneity should be taken into account as an additional substance effect (eg, the use of a location-scale trend test). Lack of balance occurs frequently in repeated toxicity studies. On the one hand sample size in negative control is increased due to the \sqrt{k} -rule, on the other hand, mortality-related loss of animals occurs (sometimes dose-dependent). Contrast tests and Dunnett's procedure can be used in the general unbalanced case, but for Williams' MCP such an algorithm does not seem to be available.

SIMULATION RESULTS

A simulation study was performed to compare power between several approaches (all procedures are level- α methods), for example, $k = 3 + 1$, $\alpha = 0.05$, $n_i = \text{const} = 10$, $N(\mu_i, \sigma^2 = 1)$, $\mu_k = 0$. Only results for selected dose-response profiles will be reported here (Tables 4–7).

Even for monotonic profiles, a uniformly powerful procedure does not exist. A closure test using a pairwise contrast reveals an average power. Adjustment for downturns causes a power loss, however, the power is higher than for the unrestricted alternative. For non-monotonic profiles the power advantage of the adjusted procedure can be seen. For variance heterogeneity the power loss of the procedures based on mean square error can be clearly seen.

EXAMPLE

Table 8 presents raw data from alkaline phosphatase measurement after six months in a repeated toxicity study in rats. It is easy to see that a possible downturn at high doses occurs and that the variances are not homogeneous.

Moreover, from historical controls it is known that the distribution of this endpoint is heavily skewed. Therefore, nonparametric

TABLE 4
Included Procedures

	Procedure	Restriction of Alternative*
DU	Dunnett (29) MCP	unrestricted
WI	Williams (14, 15) MCP: ML-estimation test	total ordered alternative
rH	Closure test on reverse Helmert contrast Fligner/Wolfe (19)	total ordered alternative
PC	Closure test based on pairwise contrasts Hothorn/Lehmacher (25)	total ordered alternative
aPC	Closure test based on pairwise contrasts adjusted for downturns	partial ordered alternative
wPC	Closure test based on pairwise contrasts using Welch tests and adjusted for downturns	partial ordered alternative
Rom	Closure test based on linear contrasts adjusted for downturns, Rom et al. (10)	partial ordered alternative
H	Closure test on Helmert contrasts	total ordered alternative

*All one-sided.

TABLE 5
Power Estimation for Monotone Dose Response Relationships*

μ_1	μ_2	μ_3	Power Estimation						
			DU	WI	Closure Tests				
					rH	PC	aPC	Rom	H
0	0	1.5	0.89	0.95	0.39	0.96	0.88	0.86	0.97
			0.02	0.02	0.05	0.05	0.05	0.05	0.05
			0.02	0.02	0.03	0.01	0.03	0.03	0.02
1.5	1.5	1.5	0.78	0.91	0.95	0.88	0.89	0.95	0.95
			0.78	0.74	0.93	0.82	0.87	0.94	0.94
			0.78	0.70	0.86	0.77	0.83	0.92	0.93
0	0.5	1.5	0.90	0.93	0.57	0.96	0.88	0.92	0.93
			0.17	0.14	0.16	0.30	0.21	0.22	0.26
			0.02	0.09	0.05	0.04	0.03	0.04	0.04
0.5	1.0	1.5	0.89	0.92	0.86	0.96	0.90	0.93	0.85
			0.56	0.69	0.61	0.71	0.62	0.66	0.57
			0.17	0.39	0.29	0.27	0.27	0.27	0.23

*Three power estimations for $\{C-,D_1,D_2,D_3\}$, $\{C-,D_1,D_2\}$ and $\{C-,D_1\}$.

TABLE 6
Power Estimation for Dose Response Relationships* with Downturns at High Doses

μ_1	μ_2	μ_3	Power Estimations						
			DU	WI	rH	PC	aPC	Rom	H
0.5	1.5	1.25	0.76	0.93	0.91	0.87	0.93	0.96	0.91
			0.90	0.93	0.82	0.95	0.90	0.92	0.89
			0.17	0.30	0.30	0.29	0.30	0.29	0.27
0.5	1.5	1.0	0.56	0.87	0.86	0.72	0.91	0.94	0.90
			0.90	0.87	0.81	0.71	0.90	0.91	0.89
			0.17	0.30	0.30	0.27	0.30	0.29	0.27
0.5	1.5	.75	0.34	0.80	0.81	0.51	0.89	0.92	0.88
			0.90	0.80	0.78	0.51	0.89	0.91	0.88
			0.17	0.30	0.30	0.22	0.30	0.29	0.27
0.5	1.5	0.5	0.17	0.69	0.74	0.30	0.89	0.91	0.88
			0.90	0.69	0.73	0.30	0.89	0.90	0.88
			0.17	0.29	0.30	0.16	0.30	0.28	0.27

*Three power estimations for $\{C-,D_1,D_2,D_3\}$, $\{C-,D_1,D_2\}$ and $\{C-,D_1\}$.

TABLE 7
Power Estimation for Linear Dose Response Relationships*
in the Case of Variance Heterogeneity

σ_1	σ_2	σ_3	Power Estimations						
			DU	WI	rH	PC	wPC	Rom	H
1	1	1	0.89	0.92	0.86	0.96	0.95	0.93	0.85
			0.56	0.69	0.61	0.71	0.70	0.66	0.57
			0.17	0.39	0.29	0.27	0.25	0.27	0.23
2	1	1	0.72	0.81	0.70	0.86	0.95	0.78	0.64
			0.33	0.59	0.43	0.53	0.70	0.43	0.41
			0.16	0.39	0.23	0.17	0.13	0.17	0.17
1	2	1	0.71	0.86	0.70	0.86	0.95	0.78	0.66
			0.40	0.53	0.45	0.51	0.40	0.47	0.37
			0.06	0.13	0.14	0.12	0.14	0.12	0.09
1	1	2	0.66	0.62	0.70	0.75	0.65	0.75	0.60
			0.34	0.40	0.42	0.46	0.51	0.41	0.29
			0.06	0.19	0.15	0.12	0.21	0.12	0.09
2	2	1	0.53	0.72	0.58	0.76	0.95	0.63	0.53
			0.27	0.46	0.34	0.38	0.40	0.34	0.33
			0.09	0.17	0.15	0.09	0.08	0.09	0.10
4	1	1	0.21	0.40	0.31	0.44	0.95	0.29	0.30
			0.05	0.23	0.20	0.13	0.70	0.08	0.24
			0.13	0.19	0.17	0.04	0.07	0.03	0.12

*Three power estimations for $\{C-,D_1,D_2,D_3\}$, $\{C-,D_1,D_2\}$ and $\{C-,D_1\}$.

tests are used, because of the small sample size in the permutative versions. Using the gold standard nonparametric dose response test, Jonckheere's (6) trend test, for the global hypothesis the p-value of 0.14 (using STATXACT, Mehta and Patel, 33) leads to a conclusion of no dose-response effect.

Alternatively, a PC procedure based on permutative U-tests is demonstrated in Table 9. This procedure reveals a significant global trend as well as a local decision on MED = 20 mg/kg.

CONCLUSIONS

Real dose-response studies in toxicology are characterized by:

1. A priori *unknown* shape of the dose-profile,
2. The *possibility of downturns* at high doses,
3. The *possibility of* (dose-dependent) *variance heterogeneity and lack of balance*, and
4. *Endpoints* other than those which are Gaussian-distributed.

TABLE 8
AP Data from a Six-Month Study

D	n_i	AP										mean	s	
0	10	17.3	23.0	22.5	10.0	21.6	19.7	17.7	20.0	17.9	24.3	19.40	4.079	
5	9	17.4	20.0	21.5	11.5	19.7	17.0	19.8	21.5	12.1		17.83	3.775	
10	11	23.9	19.2	24.1	19.1	17.4	21.9	42.1	24.0	20.7	21.9	16.9	22.84	6.881
20	10	24.2	31.3	30.7	17.4	27.9	25.7	18.1	21.9	25.9	19.6	24.27	4.959	
50	9	19.9	21.9	20.9	17.1	24.7	16.3	29.1	15.3	19.8		20.56	4.341	

TABLE 9
Decision Scheme of a PC Procedure

Hypothesis	Tests i	p_i	Criteria	Decision
Global {0,5,10,20,50}	[0,50]	0.143	$p_{\min} < \alpha/4$	trend,
	[0,20]	0.0084	0.0125?	go to the partial
	[0,10]	0.171		hypothesis
	[0,5]	0.144		{0,5,10}
Partial {0,5,10}	[0,10]	0.171	$p_{\min} < \alpha/2$	no trend and stop
	[0,5]	0.144	0.0167?	

A partial order restricted test based on any kind of two-sample test can be used. In the parametric case, the Welch t-test should be used, because in the homogenous case this is the t-test and in the heterogeneous case a correction will be performed on such comparisons where the heterogeneity occurs. This procedure can be used for closure principles of interest in toxicology: all-pair and many-to-one. This principle is based on simple univariate distribution. It guarantees methodological homogeneity within the toxicological study, for example, repeated toxicity study could use:

- Mortality-time relationship: \Rightarrow log-rank test
- Hemoglobin \Rightarrow Welch t-test
- ASAT (heavy skewed) \Rightarrow permutative u-test
- Histological findings \Rightarrow Fisher's exact test (eg, mid p-values)
- Hyperplasia \Rightarrow stratified Fisher's exact test

REFERENCES

1. Hothorn LA, Hamada Y, Rebel W, Lin K. Recommendations for Biostatistics of Repeated Toxicity Studies. *Drug Int J.* 1997;31.
2. Hauschke E. Testprinzipien bei safety Studien in der Toxikologie. In: Trampisch HJ. *Medizinische Informatik, Biometrie und Epidemiologie.* Vol. 80; München: MMV Medizin Verlag; 1995:104–107.
3. Hauschke D. Statistical proof of safety in toxicological studies. *Drug Inf. J.* 1997;31.
4. Neuhäuser M, Hothorn LA. Auswertung der Dosis-Wirkungs-Abhängigkeit des Ames Mutagenitätsassay bei direkter Kontrolle des Konsumentenrisiko. In: Trampisch HJ. *Medizinische Informatik, Biometrie und Epidemiologie.* Vol. 80; München: MMV Medizin Verlag; 1995:113–116.

5. Neuhäuser M, Hothorn LA. The control of the consumer risk in the Ames Assay. *Drug Inf J.* 1997;31.
6. Jonckheere AR. A distribution-free k-sample test against ordered alternatives. *Biometrika.* 1954;41: 133–145.
7. Marcus R, Peritz E, Gabriel KB. On closed testing procedures with special reference to ordered analysis of variance. *Biometrika.* 1976;63:655–660.
8. Budde M, Bauer P. Multiple test procedures in clinical dose finding studies. *J Am Stat Assoc.* 1989;84: 792–796.
9. Bauer P, Budde M. Multiple testing for detecting efficient dose steps. *Biometr J.* 1994;36:3–15.
10. Rom DM, Costello RJ, Connel LT. On closed test procedures for dose-response analysis. *Stat Med.* 1994;13:1583–1596.
11. Tukey JW, Ciminera JL, Heyse JF. Testing the statistical certainty of a response to increasing doses of a drug. *Biometrics.* 1985;41:295–301.
12. Barlow RE, Bartholomew DJ, Bremner JM, Brunk HD. *Statistical inference under order restrictions.* London: Wiley; 1972.
13. Robertson T, Wright FT, Dykstra RL. *Order restricted statistical inference.* New York: Wiley; 1988.
14. Williams DA. A test for differences between treatment means when several dose levels are compared with a zero dose control. *Biometrics.* 1971;27:103–117.
15. Williams DA. The comparison of several dose levels with a zero dose control. *Biometrics.* 1972;28:519–531.
16. Mehta H, Capizzi T, Oppenheimer L. Use of SAS software for trend and sequential trend test analysis. In: *Proceedings of the 9th SUGI.* SAS Institute, Cary, 1984;794–799.
17. Neuhäuser M. Trend tests for a priori unknown expected value profile. Thesis; University of Dortmund (1996) (in German).
18. Capizzi T, Survill TT, Heyse JF, Malani H. An empirical and simulated comparison of some tests for detecting progressiveness of response with increasing doses of a compound. *Biometr J.* 1992;34:275–289.
19. Fligner MA, Wolfe DA. Distribution-free tests for comparing several treatments with a control. *Stat Neerlandica.* 1982;36:119–127.
20. Bartholomew DJ. Ordered test in the analysis of variance. *Biometrika.* 1961;48:325–332.

21. Mukerjee H, Robertson T, Wright FT. Multiple contrast tests for testing against a simple tree ordering. In: Dykstra R., ed. *Advances in order restricted statistical inference*. Berlin: Springer; 1986:203–230.
22. Mukerjee H, Robertson T, Wright FT. Comparison of several treatments with a control using multiple contrasts. *J Am Stat Assoc*. 1987;82:902–910.
23. Ruberg SJ. Contrasts for identifying the minimum effective dose. *J Am Stat Assoc*. 1989;84:816–822.
24. Abelson RP, Tukey JW. Efficient utilization of non-numerical information in quantitative analysis: general theory and the case of simple order. *Ann Math Stat*. 1963;34:1347–1369.
25. Hothorn LA, Lehmacher W. A simple testing procedure control versus k treatments for one-sided ordered alternatives, with application in toxicology. *Biometrical J*. 1991;33:179–189.
26. Tamhane AC, Dunnett CW, Hochberg Y. Multiple test procedures for dose finding. *Biometrics*. 1996; 52:21–37.
27. Bechhofer RE, Dunnett CW. Multiple comparisons for orthogonal contrasts: example and table. *Technometrics*. 1982;24:213–222.
28. Maurer W, Hothorn L, Lehmacher W. Multiple comparisons in drug clinical trials and preclinical assays: a-priori ordered hypotheses. In: Vollmar J, ed. *Biometrie in der chemisch-pharmazeutischen Industrie*. Vol. 6. Stuttgart: Fischer Verlag; 1995:3–18.
29. Dunnett CW. A multiple comparison procedure for comparing several treatments with a control. *J Am Stat Assoc*. 1955;50:1096–1121.
30. Shaffer JP. Modified sequentially rejective multiple test procedures. *J Am Stat Assoc*. 1986;81:826–831.
31. Steel RGD. A multiple comparison rank sum test treatments versus control. *Biometrika*. 1959;15:560–572.
32. Ortseifen C, Hothorn LA. Multiple Vergleiche 'Kontrolle gegen k Behandlungen' bei Abweichung con den ANOVA-Annahmen. In: Vollmar J, ed. *Biometrie in der chemisch-pharmazeutischen Industrie*. Vol. 6. Stuttgart: Fischer Verlag; 1995:77–100.
33. Mehta C, Patel N. StatXact 3 for Windows, Cytel Software Corp. Cambridge;1995.