

Robust statistical evaluation of tumor incidences in long-term rodent carcinogenicity studies: the reformulated poly-k trend test

Ludwig A. Hothorn,
Im Grund 12, D-31867 Lauenau, Germany

(retired from Leibniz University Hannover)

February 14, 2022

Abstract

Mortality-adjusted tumor rates in long-term carcinogenicity rodent bioassays are commonly evaluated by means of the poly-k=3 Armitage trend test. However, this assumes exactly a linear dose-response curve and the Weibull parameter of k=3 for all tumor sites. These unrealistic assumptions can be circumvented by multiple testing across multiple possible dose-response shapes, multiple Weibull parameters, multiple effect sizes, multiple correlated tumors as well as pairwise and trend tests using the multiple marginal models approach. Based on data examples, different multiple tests are demonstrated using the CRAN R packages *multcomp*, *tukeytrend*, *coin*, *MCPAN* and *multfisher*.

1 The problem

In NTP-designed studies, several site-specific tumor incidences in a negative control and a few dose groups are compared for a possible increasing trend, e.g. using the Armitage trend test (CA) [3]. A positive finding is when any tumor type in either males or females $p^{CA} < 0.05$ - this appears to be a simple and reproducible procedure.

But it is unfortunately not the case: i) tumor development and mortality are interdependent in a complex way, ii) the different tumors found according to an a priori list or spontaneously (and their summaries (e.g., adenomas of the endocrine system) can be modeled as correlated primary endpoints, iii) in some evaluations , pairwise comparisons are performed to control using Fisher tests in addition to the CA-test, iv) the CA test is defined for linear dose-response dependencies (shows highest power there, but not for e.g. plateau-shape profiles), v) either one-or two-tailed trend tests are reported, this has a significant impact on the false negative rate), vi) $p^{CA} < 0.05$ vs $p^{CA} < 0.01$ relevance criteria for common vs. rare tumors. The specific choice of these six criteria significantly influences the ratio of false positive (f+) to false negative (f-) error rates, the actual and primary issue in these safety studies, where additional effects being discussed. In the following chapter, these criteria are quantified individually from a statistical point of view, with a further underlying issue being various sources of multiplicity. The third underlying problem is the a priori restricted vs. unrestricted formulation of the alternative hypothesis. Based on this balancing of interests, a proposal of an evaluation procedure is made using the multiple marginal tests approach [15] in sub-chapter 2.5.. The idea is not to use only one, a-priori (or per guideline) defined model, but to use several, well-justified models, in order to be able to represent several conceivable data situations. In a maxT test, the best-fitting model then determines the p-value, with the conservativeness price of the multiple models being limited by estimating their correlations. A flexible concept, but with two disadvantages: it is only asymptotically valid, i.e. for infinitely large n_i , and so far it does not adequately account for the specific discreteness

and sparseness of tumor findings.

2 A the reformulated poly-k trend test

2.1 Choosing the poly-k parameter k

The relationship between tumor development and mortality is complex: on the one hand, animals that die early cannot develop tumors, on the other hand, tumors can preliminary diagnosed in animals that die or were sacrificed, particularly early. Therefore, the analysis of crude tumor incidences may be biased and should generally be avoided. A simple method of mortality adjustment is the poly-k adjustment [5], where the difficult-to-achieve cause-of-death information is not necessary. Possible individual-specific mortality differences are considered by individual weights $w_{ij} = (t_{ij}/t_{max})^k$ (t_{ij} ... time of death of animal j in dose i). The weights result in adjusted sample sizes $n_i^* = \sum_{j=1}^{n_i} w_{ij}$ and adjusted proportions $p_i^* = y_i/n_i^*$ instead of the randomized samples sizes n_i and crude tumor proportions $p_i = y_i/n_i$. The shape parameter k in a range between $k = 1, \dots, 5$ reflects different Weibull hazard functions for cumulative tumor incidences over time. The question arises, which k should be taken specifically (i.e. per tumor site)? The majority of published evaluations take exactly $k = 3$, other studies are performed in parallel for $k=3$ and $k=6$ (each to the level α). Here we propose a joint $\max T^{mmm}$ test over all values of k , yielding the p-value for the best-fitting k value in each case. Since the six models are highly correlated, the penalty of the multiplicity correction appears to be reasonable.

As an example the incidence of skin fibroma in the bioassay of methyleugenol using rats was used [14], available in the package MCPAN [18]. Table 1 reveals the smallest p-value for $\text{poly}k = 5$ (in bold), considerably smaller than the commonly used $k = 3$ model (even if it had been chosen a priori alone). The magnitude of multiplicity adjustment is illustrated by comparing the marginal with the adjusted p-value for $k=5$ of $p = 0.0157$ and $p = 0.0205$, i.e., if one had known a priori that this was the best assumption. The related R-code is available in the Appendix.

k	estimate	s.e.	t	p^{adj}	alone
1	0.0006	0.0005	1.3318	0.1097	
2	0.0008	0.0005	1.5619	0.0724	
3	0.0010	0.0006	1.7740	0.0471	0.0391
4	0.0012	0.0006	1.9701	0.0310	
5	0.0013	0.0006	2.1520	0.0205	0.0157
6	0.0010	0.0006	1.7740	0.0476	

Table 1: Adjusted p-values for 6 jointly analysed poly- $k = 1, \dots, 6$ tests

This method seems to me recommendable if one imagines that a specific k for the different tumor sites per sex and species is optimal.

2.2 Modeling dose as quantitative covariate and qualitative factor, sensitive to more than just linear shapes

The original work of poly-k test [5], is based on the CA trend test [4], formulated precisely for a linear regression model in the generalized linear model (GLM). Given the numerous tumor sites alone, the shape of the dose-response function is not an assumption but a study result. Thus, the use of a trend test is indicated, which guarantees an acceptable power for several shapes, precisely for profiles with plateau shape up to downturns at high dose(s). Already Tukey proposed a maxT test for arithmetic, ordinal and logarithmic (because of D0 precisely arithmetic-logarithmic) dose scores [20], which is now available extended for the glm based on the mmm approach [19]. This approach should be extended by the Williams contrast test [21] (and thus the dose is considered qualitatively and quantitatively, jointly), since this highly-multiple test then detects linear shapes as well as shapes with plateau and

even downturns with some power [10]. Although these 4 models are highly correlated, one must accept a multiplicity price. This will be demonstrated by means of the methyleugenol bioassay example (and k=5):

Model	estimate	se	t	p^{adj}
Regression: arithmetic scores	0.0013	0.0006	2.1520	0.0357
Regression: ordinal scores	0.0772	0.0272	2.8361	0.0058
Regression: ari-logarithmic scores	0.1103	0.0392	2.8122	0.0062
Williams contrast: $p_3 - p_0$	0.1640	0.0808	2.0286	0.0472
Williams contrast: $(p_3 + p_2)/2 - p_0$	0.1837	0.0585	3.1377	0.0022
Williams contrast: $(p_3 + p_2 + p_1)/3 - p_0$	0.1953	0.0497	3.9303	0.0001

Table 2: Tukey-Williams trend test for poly-k=5

If one knew a-priori the pooled Williams contrast would be the best-fitting model, then its marginal p-value of only 0.000039 (compared to 0.000109 illustrates the extent of multiplicity- but remember the p-value for an a priori chosen linear CA trend test (k=5) of 0.0157. I.e. the improvement can be substantial, depending on data.

2.3 Considering multiple tumor sites simultaneously

One can consider selected or, in extreme cases, all tumor sites as multiple correlated endpoints and thus test them jointly with a maxT test. In clear contrast to issues i and ii), this method is inherently conservative, i.e., a power drop compared to the marginal independent analysis of the individual tumors (each at level α) is the consequence. In addition to adherence to the familywise error rate (FWER) (a concept that may not be immediately obvious to toxicologists), the advantage is the joint interpretation of different tumor sites. Once again, the mmm approach is suitable for formulating a double max(max)T test: across multiple tumors and across multiple contrasts or multiple regression models, jointly. Three approaches can be used: i) conditional inference using the package coin [12] (see an example in [7]), ii) multiple contrasts using the package multcomp [11] a further example in [7] and, iii) multiple regression models in the package tukeytrend [19] (see an example in [10]). An example using Williams Trend tests for five selected crude liver tumors (hepatoblastoma, hepatocellular adenoma, hepatocellular carcinoma, hepatocholangiocarcinoma (t24,t27,t29,t30)) in female mice from dataset NTP-TR491 [2] is demonstrated below (see the R-code in the Appendix):

Tumor site	Contrast	estimate	se	t	p
Hepatoblastoma	$p_3 - p_0$	2.7279	1.0620	2.5685	0.0093
	$(p_3 + p_2)/2 - p_0$	2.7279	1.0620	2.5685	0.0093
	$(p_3 + p_2 + p_1)/3 - p_0$	2.5089	1.0708	2.3430	0.0168
Hepatocellular adenoma	$p_3 - p_0$	2.1160	0.4662	4.5393	0.000004
	$(p_3 + p_2)/2 - p_0$	2.3893	0.4744	5.0364	0.0000002
	$(p_3 + p_2 + p_1)/3 - p_0$	2.5155	0.4788	5.2542	0.00000008
Hepatocellular carcinoma	$p_3 - p_0$	5.0304	1.0593	4.7487	0.000001
	$p_3 - p_0$	4.6362	1.0537	4.4000	0.000008
	$(p_3 + p_2 + p_1)/3 - p_0$	4.3758	1.0515	4.1613	0.00005
Hepatocholangiocarcinoma	$p_3 - p_0$	1.1386	1.1719	0.9716	0.2330
	$p_3 - p_0$	0.5693	1.2020	0.4737	0.4119
	$(p_3 + p_2 + p_1)/3 - p_0$	0.3795	1.1654	0.3257	0.4713

Table 3: Multiple tumors Williams trend test

In the joint observation, 4 of the 5 liver tumors show a clear increase in incidences, with the hepatocellular adenoma showing the clearest increase in the shape of a plateau (R-code see in the Appendix).

2.4 Choosing effect size: RD, RR, OR

The usual approach is to define the effect size a priori, depending on the design, the test, the interpretation, and so on. When using the p-value, the effect size is often hidden. Mostly the different effect sizes lead to a very different magnitude of the alternative. One can choose some meaningful effect sizes and formulate a maxT- test using the mmm approach [10], e.g. for proportions the risk difference (RD), risk ratio (RR) and odds ratio (OR). When $p_0 = 0$ is possible or small sample sizes occur, the add-1 adjustment can be recommended [17, 1]. As an example, the poly-k=3 Tukey-type trend tests for hepatoblastoma in female mice [2] is demonstrated below (see the R-code in the Appendix):

Effect size	Dose score	estimate	se	t	p
Odds ratio	arithmetic	0.8458	0.2249	3.7607	0.0003
	ordinal	0.8458	0.2249	3.7607	0.0002
	ari.-logarithmic	1.5428	0.4112	3.7516	0.0003
Risk difference	arithmetic	0.1321	0.0209	6.3313	< 0.00001
	ordinal	0.1321	0.0209	6.3313	< 0.00001
	ari.-logarithmic	0.2064	0.0327	6.3140	< 0.00001
Risk ratio	arithmetic	0.6215	0.1715	3.6232	0.0004
	ordinal	0.6215	0.1715	3.6232	0.0004
	ari.-logarithmic	1.1766	0.3260	3.6089	0.0003

Table 4: Multiple effect sizes for poly-k=5 Tukey trend tests

In this specific example, risk difference as an effect measure shows the lowest p-value.

2.5 Pairwise tests vs. trend test

Sometimes trend tests and pairwise tests (control vs. doses) are performed, presumably to be robust against downturns at high doses. If these elementary tests are then performed each at the α level, the false positive rate increases significantly. Both principles can also be tested simultaneously to keep the FWER [7]. The first approach based on the conditional inference using the package coin [12]. For the hepatoblastoma in female mice example, the contrasts for Williams and Dunnett test [13] are written explicitly (see the R-code in the Appendix) whereas the first Williams and first Dunnett contrast, namely $p_3 - p_0$ are double and therefore complete correlated (denoted as double):

Type	Contrast	p^{adj}	Comment
Williams	$p_3 - p_0$	0.0027	
	$(p_3 + p_2)/2 - p_0$	0.00041	
	$(p_3 + p_2 + p_1)/3 - p_0$	0.00163	
Dunnett	$p_3 - p_0$	0.0027	double
	$p_2 - p_0$	0.0027	
	$p_1 - p_0$	0.106	

Table 5: Permutative Dunnett-Williams test for poly-k=5

Not surprisingly in this data example, the smallest p-value for the partial plateau profile of $(p_3 + p_2)/2 - p_0$. Alternatively, one can take the asymptotic Dunnett-Tukey approach, demonstrated for the skin fibroma example using k=5 (see chapter 2.1):

The magnitude of the increase in the false positive rate can be estimated from the 3 p-values for the Dunnett contrast $p_1 - p_0$: i) adjusted for both Tukey and Dunnett test $p^{adj} = 0.0065$, the Dunnett test alone $p^{Dunnett} = 0.0057$ and the unadjusted pairwise contrast $p^{unadj} = 0.0019$ (notice, for asymptotic tests).

Model	Dose scores resp. contrast	estimate	se	t	p^{adj}	$p^{Dunnett}$	p^{unadj}
Tukey	arithmetic	0.0013	0.0006	2.1520	0.0535		
	ordinal	0.0772	0.0272	2.8361	0.0084		
	ari.-logarithmic	0.1103	0.0392	2.8122	0.0090		
Dunnett	$p_1 - p_0$	0.2186	0.0755	2.8945	0.0065	0.0057	0.0019
	$p_2 - p_0$	0.2033	0.0758	2.6817	0.0138	0.0109	0.0037
	$p_3 - p_0$	0.1640	0.0808	2.0286	0.0700	0.0613	0.0213

Table 6: Tukey-Dunnett trend test for poly-k=5

2.6 Consideration of covariates

If there could be a relationship between tumor progression and body or organ weights, a covariance analysis in the GLM would be useful. Here, too, several models can be tested simultaneously, e.g. without covariate, with exactly one or all covariates, see e.g. the relative organ weight issue [7].

2.7 The impact of two vs. one-tailed tests on f+ rate

Two-sided testing is not problem adequate , as only increasing tumor incidences are of interest and unnecessarily conservative. This is demonstrated using the Tukey-Dunnett test from chapter 2.5):

Model	Dose scores resp. contrast	estimate	se	t	$p_{1-sided}^{adj}$	$p_{2-sided}^{adj}$
Tukey	arithmetic	0.0013	0.0006	2.1520	0.054	0.11
	ordinal	0.0772	0.0272	2.8361	0.008	0.017
	ari.-logarithmic	0.1103	0.0392	2.8122	0.009	0.018
Dunnett	$p_1 - p_0$	0.2186	0.0755	2.8945	0.007	0.014
	$p_2 - p_0$	0.2033	0.0758	2.6817	0.013	0.026
	$p_3 - p_0$	0.1640	0.0808	2.0286	0.0700	0.14

Table 7: Tukey-Dunnett trend test for poly-k=5: 1-sided vs. 2 sided hypotheses

2.8 The availability of simultaneous confidence interval

We have been in a p-value centered interpretation. This has advantages and disadvantages. Therefore, the availability of simultaneous confidence intervals, as compatible as possible with the adjusted p-values, is an advantage - and this is exactly what the mmm-approach offers. As an example the two-sided confidence limits for the six poly-k parameters ($k = 1, 2, 3, 4, 5, 6$) and the three tumor sites (hepatoblastoma, hepatocellular adenoma, hepatocellular carcinoma) can be estimated with the code in the Appendix.

2.9 An approach which includes many models simultaneously

In the following, we demonstrate that even multiple models can be considered simultaneously in one approach: i) multiple k, ii) multiple regression models, and iii) multiple tumor sites. Not that I would recommend such a complex approach for routine re-evaluation. The point here is to illustrate the trade-off between the broadest possible alternatives and still permissible conservatism by means of an example using three tumor sites (hepatoblastoma, hepatocellular adenoma, hepatocellular carcinoma; t24,t27,t29), 2 effect sizes (odds ratio, risk difference; OR, RD), 6 poly-k parameter ($k = 1, 2, 3, 4, 5, 6$) and 3 regression models (arithmetic, ordinal, ari.logarithmic), which results in the crazy number of 108 models to be tested simultaneously:

No	Model	Tumor site	estimate	se	t	p^{adj}	p^{unadj}
1	tuOR241.glm.t24.doseari: doseari	T24	0.7296	0.2149	3.3950	0.0072	0.0008
2	tuOR241.glm.t24.doseord: doseord		0.7296	0.2149	3.3950	0.0075	0.0008
3	tuOR241.glm.t24.dosearilog: dosearilog		1.3577	0.3955	3.4331	0.0060	0.0007
4	tuOR242.glm.t24.doseari: doseari		0.7601	0.2172	3.4997	0.0046	0.0006
5	tuOR242.glm.t24.doseord: doseord		0.7601	0.2172	3.4997	0.0049	0.0006
6	tuOR242.glm.t24.dosearilog: dosearilog		1.4072	0.3991	3.5259	0.0044	0.0005
7	tuOR243.glm.t24.doseari: doseari		0.7896	0.2196	3.5952	0.0037	0.0004
8	tuOR243.glm.t24.doseord: doseord		0.7896	0.2196	3.5952	0.0033	0.0004
9	tuOR243.glm.t24.dosearilog: dosearilog		1.4544	0.4030	3.6094	0.0035	0.0004
10	tuOR244.glm.t24.doseari: doseari		0.8181	0.2222	3.6818	0.0025	0.0003
11	tuOR244.glm.t24.doseord: doseord		0.8181	0.2222	3.6818	0.0029	0.0003
12	tuOR244.glm.t24.dosearilog: dosearilog		1.4995	0.4070	3.6842	0.0023	0.0003
13	tuOR245.glm.t24.doseari: doseari		0.8458	0.2249	3.7607	0.0021	0.0002
14	tuOR245.glm.t24.doseord: doseord		0.8458	0.2249	3.7607	0.0018	0.0002
15	tuOR245.glm.t24.dosearilog: dosearilog		1.5428	0.4112	3.7516	0.0023	0.0002
16	tuOR246.glm.t24.doseari: doseari		0.8727	0.2277	3.8330	0.0015	0.0002
17	tuOR246.glm.t24.doseord: doseord		0.8727	0.2277	3.8330	0.0017	0.0002
18	tuOR246.glm.t24.dosearilog: dosearilog		1.5845	0.4156	3.8128	0.0017	0.0002
19	tuRD241.glm.t24.doseari: doseari		0.1011	0.0168	6.0207	0.0000	0.0000
20	tuRD241.glm.t24.doseord: doseord		0.1011	0.0168	6.0207	0.0000	0.0000
21	tuRD241.glm.t24.dosearilog: dosearilog		0.1580	0.0263	6.0140	0.0000	0.0000
22	tuRD242.glm.t24.doseari: doseari		0.1097	0.0180	6.1018	0.0000	0.0000
23	tuRD242.glm.t24.doseord: doseord		0.1097	0.0180	6.1018	0.0000	0.0000
24	tuRD242.glm.t24.dosearilog: dosearilog		0.1716	0.0282	6.0931	0.0000	0.0000
25	tuRD243.glm.t24.doseari: doseari		0.1178	0.0191	6.1806	0.0000	0.0000
26	tuRD243.glm.t24.doseord: doseord		0.1178	0.0191	6.1806	0.0000	0.0000
27	tuRD243.glm.t24.dosearilog: dosearilog		0.1841	0.0298	6.1694	0.0000	0.0000
28	tuRD244.glm.t24.doseari: doseari		0.1252	0.0200	6.2570	0.0000	0.0000
29	tuRD244.glm.t24.doseord: doseord		0.1252	0.0200	6.2570	0.0000	0.0000
30	tuRD244.glm.t24.dosearilog: dosearilog		0.1957	0.0313	6.2430	0.0000	0.0000
31	tuRD245.glm.t24.doseari: doseari		0.1321	0.0209	6.3313	0.0000	0.0000
32	tuRD245.glm.t24.doseord: doseord		0.1321	0.0209	6.3313	0.0000	0.0000
33	tuRD245.glm.t24.dosearilog: dosearilog		0.2064	0.0327	6.3140	0.0000	0.0000
34	tuRD246.glm.t24.doseari: doseari		0.1385	0.0216	6.4035	0.0000	0.0000
35	tuRD246.glm.t24.doseord: doseord		0.1385	0.0216	6.4035	0.0000	0.0000
36	tuRD246.glm.t24.dosearilog: dosearilog		0.2164	0.0339	6.3824	0.0000	0.0000
37	tuOR271.glm.t27.doseari: doseari	T27	0.6716	0.1482	4.5304	0.0001	0.0000
38	tuOR271.glm.t27.doseord: doseord		0.6716	0.1482	4.5304	0.0001	0.0000
39	tuOR271.glm.t27.dosearilog: dosearilog		1.1888	0.2412	4.9296	0.0000	0.0000
40	tuOR272.glm.t27.doseari: doseari		0.7157	0.1543	4.6393	0.0000	0.0000
41	tuOR272.glm.t27.doseord: doseord		0.7157	0.1543	4.6393	0.0001	0.0000
42	tuOR272.glm.t27.dosearilog: dosearilog		1.2559	0.2498	5.0285	0.0000	0.0000
43	tuOR273.glm.t27.doseari: doseari		0.7546	0.1598	4.7226	0.0000	0.0000
44	tuOR273.glm.t27.doseord: doseord		0.7546	0.1598	4.7226	0.0000	0.0000
45	tuOR273.glm.t27.dosearilog: dosearilog		1.3145	0.2575	5.1042	0.0000	0.0000
46	tuOR274.glm.t27.doseari: doseari		0.7888	0.1648	4.7853	0.0000	0.0000
47	tuOR274.glm.t27.doseord: doseord		0.7888	0.1648	4.7853	0.0001	0.0000
48	tuOR274.glm.t27.dosearilog: dosearilog		1.3660	0.2647	5.1606	0.0000	0.0000
49	tuOR275.glm.t27.doseari: doseari		0.8196	0.1696	4.8337	0.0000	0.0000
50	tuOR275.glm.t27.doseord: doseord		0.8196	0.1696	4.8337	0.0000	0.0000
51	tuOR275.glm.t27.dosearilog: dosearilog		1.4119	0.2714	5.2031	0.0000	0.0000
52	tuRD276.glm.t27.doseari: doseari		0.8477	0.1740	4.8722	0.0001	0.0000
53	tuRD276.glm.t27.doseord: doseord		0.8477	0.1740	4.8722	0.0001	0.0000
54	tuRD276.glm.t27.dosearilog: dosearilog		1.4539	0.2776	5.2369	0.0000	0.0000
55	tuRD271.glm.t27.doseari: doseari		0.1344	0.0291	4.6249	0.0001	0.0000
56	tuRD271.glm.t27.doseord: doseord		0.1344	0.0291	4.6249	0.0001	0.0000
57	tuRD271.glm.t27.dosearilog: dosearilog		0.2499	0.0475	5.2617	0.0000	0.0000
58	tuRD272.glm.t27.doseari: doseari		0.1366	0.0291	4.6881	0.0000	0.0000
59	tuRD272.glm.t27.doseord: doseord		0.1366	0.0291	4.6881	0.0000	0.0000
60	tuRD272.glm.t27.dosearilog: dosearilog		0.2537	0.0478	5.3081	0.0000	0.0000
61	tuRD273.glm.t27.doseari: doseari		0.1381	0.0292	4.7351	0.0001	0.0000
62	tuRD273.glm.t27.doseord: doseord		0.1381	0.0292	4.7351	0.0001	0.0000
63	tuRD273.glm.t27.dosearilog: dosearilog		0.2562	0.0480	5.3399	0.0000	0.0000
64	tuRD274.glm.t27.doseari: doseari		0.1390	0.0292	4.7665	0.0000	0.0000
65	tuRD274.glm.t27.doseord: doseord		0.1390	0.0292	4.7665	0.0000	0.0000
66	tuRD274.glm.t27.dosearilog: dosearilog		0.2578	0.0481	5.3572	0.0000	0.0000
67	tuRD275.glm.t27.doseari: doseari		0.1394	0.0291	4.7874	0.0000	0.0000
68	tuRD275.glm.t27.doseord: doseord		0.1394	0.0291	4.7874	0.0000	0.0000
69	tuRD275.glm.t27.dosearilog: dosearilog		0.2586	0.0482	5.3650	0.0000	0.0000
70	tuRD276.glm.t27.doseari: doseari		0.1397	0.0291	4.8021	0.0000	0.0000
71	tuRD276.glm.t27.doseord: doseord		0.1397	0.0291	4.8021	0.0001	0.0000
72	tuRD276.glm.t27.dosearilog: dosearilog		0.2591	0.0483	5.3677	0.0000	0.0000
73	tuRD291.glm.t29.doseari: doseari	T29	0.2845	0.0150	18.9618	0.0000	0.0000
74	tuRD291.glm.t29.doseord: doseord		0.2845	0.0150	18.9618	0.0000	0.0000
75	tuRD291.glm.t29.dosearilog: dosearilog		0.4625	0.0252	18.3823	0.0000	0.0000
76	tuRD292.glm.t29.doseari: doseari		0.2919	0.0145	20.1452	0.0000	0.0000
77	tuRD292.glm.t29.doseord: doseord		0.2919	0.0145	20.1452	0.0000	0.0000
78	tuRD292.glm.t29.dosearilog: dosearilog		0.4772	0.0244	19.5901	0.0000	0.0000
79	tuRD293.glm.t29.doseari: doseari		0.2970	0.0141	21.0489	0.0000	0.0000
80	tuRD293.glm.t29.doseord: doseord		0.2970	0.0141	21.0489	0.0000	0.0000
81	tuRD293.glm.t29.dosearilog: dosearilog		0.4884	0.0235	20.7453	0.0000	0.0000
82	tuRD294.glm.t29.doseari: doseari		0.3003	0.0140	21.4996	0.0000	0.0000
83	tuRD294.glm.t29.doseord: doseord		0.4970	0.0228	21.8414	0.0000	0.0000
84	tuRD295.glm.t29.dosearilog: dosearilog		0.3020	0.0141	21.4471	0.0000	0.0000
85	tuRD295.glm.t29.doseari: doseari		0.3020	0.0141	21.4471	0.0000	0.0000
86	tuRD295.glm.t29.doseord: doseord		0.5038	0.0220	22.8779	0.0000	0.0000
88	tuRD296.glm.t29.doseari: doseari		0.3025	0.0144	21.0292	0.0000	0.0000
89	tuRD296.glm.t29.doseord: doseord		0.3025	0.0144	21.0292	0.0000	0.0000
90	tuRD296.glm.t29.dosearilog: dosearilog		0.5092	0.0213	23.8581	0.0000	0.0000
91	tuRD291.glm.t29.doseari: doseari		1.2399	0.1815	6.8304	0.0000	0.0000
92	tuRD291.glm.t29.doseord: doseord		1.2399	0.1815	6.8304	0.0000	0.0000
93	tuRD291.glm.t29.dosearilog: dosearilog		2.0831	0.3005	6.9311	0.0000	0.0000
94	tuRD292.glm.t29.doseari: doseari		1.3034	0.1902	6.8518	0.0000	0.0000
95	tuRD292.glm.t29.doseord: doseord		1.3034	0.1902	6.8518	0.0000	0.0000
96	tuRD292.glm.t29.dosearilog: dosearilog		2.1784	0.3121	6.9803	0.0000	0.0000
97	tuRD293.glm.t29.doseari: doseari		1.3605	0.1985	6.8541	0.0000	0.0000
98	tuRD293.glm.t29.doseord: doseord		1.3605	0.1985	6.8541	0.0000	0.0000
99	tuRD293.glm.t29.dosearilog: dosearilog		2.2634	0.3231	7.0056	0.0000	0.0000
100	tuRD294.glm.t29.doseari: doseari		1.4120	0.2064	6.8427	0.0000	0.0000
101	tuRD294.glm.t29.doseord: doseord		1.4120	0.2064	6.8427	0.0000	0.0000
102	tuRD294.glm.t29.dosearilog: dosearilog		2.3394	0.3336	7.0127	0.0000	0.0000
103	tuRD295.glm.t29.doseari: doseari		1.4588	0.2138	6.8225	0.0000	0.0000
104	tuRD295.glm.t29.doseord: doseord		1.4588	0.2138	6.8225	0.0000	0.0000
105	tuRD295.glm.t29.dosearilog: dosearilog		2.4078	0.3436	7.0072	0.0000	0.0000
106	tuRD296.glm.t29.doseari: doseari		1.5016	0.2209	6.7973	0.0000	0.0000
107	tuRD296.glm.t29.doseord: doseord		1.5016	0.2209	6.7973	0.0000	0.0000
108	tuRD296.glm.t29.dosearilog: dosearilog		2.4701	0.3532	6.9935	0.0000	0.0000

Table 8: Considering multiple tumors, poly-k parameters, effect sizes and regression models jointly

However, the multiplicity penalty is not too big. E.g., for the hepatocellular adenoma (T27), the standard procedure (poly-k=3 CA test) yields an unadjusted p-value of 0.000054. In comparison, the adjusted p-value of 0.000068 is only slightly higher, compared to the Bonferroni adjusted p-value of

0.0059.

2.10 Further modeling aspects

There are several conditions that can be selected, which influence the ratio f+/f-. According to the guidelines, not exactly 1 study per se is evaluated, but several: females and males, rats and mice. The standard is currently to analyze them separately, each at the alpha level. From a statistical point of view, this is not mandatory. For example, one can analyze tumors of the liver for a factorial plant with primary factor dose and secondary factor sex. One could also analyze this for selected tumors across rats and mice in this way. For some important substances, there are even several studies from several laboratories with widely varying dosages. These can be evaluated e.g. in a mixed model [9]. Furthermore, the spontaneous rate of the current control p_0 as well as the historical control has a significant influence - poly-k approaches should be made available for this.

An interesting aspect is the influence of the unbalanced design on f+ rate. First, from the perspective of duplicate controls, and second, lower n_i due to group-specific mortality patterns [8].

3 Taking discreteness into account

In multiple testing of proportions, when using asymptotic tests, the level cannot be exhausted, which occurs due to the discreteness of the data (especially when $p_0 = 0$). One can use approximations, like add1 [17], or conditional or unconditional tests which use the FWER optimally. Only a few algorithms are currently available for this, and mostly for the two-sample case. With the help of the closed testing procedure, multiple tests can be formulated. In the following, two exact methods are empirically compared with the asymptotic approach on the basis of the mice dataset for the three crude proportions for hepatoblastoma, hepatocellular adenoma, hepatocellular carcinoma.

First, the condition coin-based Dunnett-type test, i.e. a conditional exact max(max)T-test [12], second the exact multivariate Fisher test [16] Bonferroni-adjusted for the 3 two-sample designs $Cvs.D_j$ and the asymptotic add-1 approach [17].

Tumor site	Contrast	$p_{coin-max(max)}^{adj}$	$p_{multiFisher}^{adj}$	$p_{add-1asymptotic}^{adj}$
Hepatoblastoma	$p_1 - p_0$	0.266	0.040	0.85
	$p_2 - p_0$	0.0067	7.51e-10	0.48
	$p_3 - p_0$	0.0067	4.57e-09	0.48
Hepatocellular adenoma	$p_1 - p_0$	1.39e-09	0.00079	0.0065
	$p_2 - p_0$	5.11e-09	2.82e-09	0.0088
	$p_3 - p_0$	5.72e-06	1.60e-10	0.046
Hepatocellular carcinoma	$p_1 - p_0$	6.69e-06	0.0024	0.022
	$p_2 - p_0$	2.41e-08	8.46e-09	0.0044
	$p_3 - p_0$	8.00e-14	4.81e-10	0.00019

Table 9: Crude proportions: exact tests vs. asymptotic test

On the one hand, the exact estimation of very small p-values is imprecise, on the other hand, from a toxicological point of view it does not matter whether a p-value is $8.00e-14$ or $4.81e-10$. But the potential of exact tests in this 2 times k panel data seems to be considerable. However, currently even low-dimensional models are extremely computationally expensive, so much so that routine evaluation is rather difficult.

4 Summary

The current standard for evaluating mortality-adjusted tumor rates in long-term carcinogenicity studies is to use the poly-k=3 Armitage trend test, per tumor separately and each at the α level. Assuming

that a trend holds exactly on a linear alternative and exactly for the Weibull distribution parameter $k = 3$ for all tumor sites seems quite unrealistic. Here, we propose a method of multiple testing based on the multiple marginal models approach with control of FWER that jointly models different dose-response models (where dose is modeled both quantitatively and qualitatively), different Weibull parameters k , different effect sizes, multiple covariates, multiple correlated tumors as well as pairwise and trend tests. Of course, this approach is more conservative than the independent marginal tests, which result in a higher false positive rate. But not extremely conservative (like a Bonferroni test), because the correlations between all marginal tests are taken into account. The basic problem is that a not precisely formulated question ("does a trend exist for this tumor?") is answered with a multitude of possible alternatives. This is robust, but conservative. This new approach can be evaluated numerically with existing CRAN R packages. The main limitation is that it is defined for asymptotic tests, but the real n_i can be very finite, especially since the table data can be very sparse with the statistically challenging $p_0 = 0$ case. A way out could be the use of exact tests, which then make exploitation of the level better. A sketch for this has been implemented, but additional work is needed for multiple testing for sparse discrete data, e.g. following [6].

Although the level of complexity outlined above will certainly frighten toxicologists (and is certainly not suitable for routine use to this extent)-the current poly- $k=3$ Armitage test carries the risk of unacceptably high false negative rates when analyzing one tumor, but also high false positive rates given > 50 tumors considered.

References

- [1] A. Agresti and B. Caffo. Simple and effective confidence intervals for proportions and differences of proportions result from adding two successes and two failures. *American Statistician*, 54(4):280–288, 2000.
- [2] anonymous. Toxicology and carcinogenesis studies of methyleugenol in f344/n rats and b6c3f1 mice. technical report, 2000, technicalreport 491. *National Toxicology Program*, 2000.
- [3] anonymous. Guidelines for carcinogenrisk assessment. *Risk assessment forum. EPA/630/P-03/001B*, 2005.
- [4] P. Armitage. Tests for linear trends in proportions and frequencies. *Biometrics*, 11(3):375–386, 1955.
- [5] A.J. Bailer and C. J. Portier. Effects of treatment-induced mortality and tumor-induced mortality on tests for carcinogenicity in small samples. *Biometrics*, 44(2):417–431, June 1988.
- [6] L. He and J.F. Heyse. Improved power of familywise error rate procedures for discrete data under dependency. *Biometrical Journal*, 61(1):101–114, 2019.
- [7] L.A. Hothorn. *Statistics in Toxicology- using R*. Chapman Hall, 2016.
- [8] L.A. Hothorn. Evaluation of tumors in long-term carcinogenicity studies: false positive vs. false negative rate. *Internal Report*, 2022.
- [9] L.A. Hothorn. Multiple comparison procedures renewed. an overview of the dunnett s test and its application: using cran r packages. <https://zenodo.org/DOI 10.5281/zenodo.5913372>, 2022.
- [10] L.A. Hothorn, F. Schaarschmidt, and C. Ritz. The maxt-test based on multiple marginal models- a vignette to library(multcomp). <https://zenodo.org/record/4421243>, 2021.
- [11] T. Hothorn, F. Bretz, and P. Westfall. Simultaneous inference in general parametric models. *Biometrical J*, 50(3):346–363, 2008.
- [12] Torsten Hothorn, Kurt Hornik, Mark A Van De Wiel, and Achim Zeileis. A lego system for conditional inference. *The American Statistician*, 60(3):257–263, 2006.
- [13] T. Jaki and L. A. Hothorn. Statistical evaluation of toxicological assays: Dunnett or Williams test-take both. *Archives of Toxicology*, 87(11):1901–1910, November 2013.
- [14] S. D. Peddada, G. E. Dinse, and J. K. Haseman. A survival-adjusted quantal response test for comparing tumour incidence rates. *Journal of the Royal Statistical Society Series C-applied Statistics*, vol. 54, pp. 51–61, 2005, 54:51–61, 2005.
- [15] C. B. Pipper, C. Ritz, and H. Bisgaard. A versatile method for confirmatory evaluation of the effects of a covariate in multiple models. *Journal of the Royal Statistical Society Series C-Applied Statistics*, 61:315–326, 2012.

- [16] R. Ristl, D. Xi, E. Glimm, and M. Posch. Optimal exact tests for multiple binary endpoints. *Computational Statistics & Data Analysis*, 122:1–17, 2018.
- [17] F. Schaarschmidt, E. Biesheuvel, and L. A. Hothorn. Asymptotic simultaneous confidence intervals for many-to-one comparisons of binary proportions in randomized clinical trials. *Journal of Biopharmaceutical Statistics*, 19(2):292–310, 2009.
- [18] Frank Schaarschmidt, Daniel Gerhard, and Martin Sill. *MCPAN: Multiple comparisons using normal approximation*, 2013. R package version 1.1-15.
- [19] Frank Schaarschmidt, Christian Ritz, and Ludwig A Hothorn. The tukey trend test: Multiplicity adjustment using multiple marginal models. *Biometrics*, 2021.
- [20] J. W. Tukey, J. L. Cimineri, and J. F. Heyse. Testing the statistical certainty of a response to increasing doses of a drug. *Biometrics*, 41(1):295–301, 1985.
- [21] D.A. Williams. A test for differences between treatment means when several dose levels are compared with a zero dose control. *Biometrics*, 27(1):103–117, 1971.

5 Appendix: R-code

5.1 R-code: poly k=1,...,6 trend tests

```
#####
# all k CA test
library(MCPAN)
data("methyl", package="MCPAN")
data(methyl)
me <- methyl
me$weightpoly1 <- 1; me$weightpoly2 <- 1; me$weightpoly3 <- 1
me$weightpoly4 <- 1; me$weightpoly5 <- 1; me$weightpoly6 <- 1
# Animals without tumor at time of death get corrected sample size
wt0 <- which(me$tumour == 0)
me$weightpoly1[wt0] <- (me$death[wt0]/max(me$death))^1
me$weightpoly2[wt0] <- (me$death[wt0]/max(me$death))^2
me$weightpoly3[wt0] <- (me$death[wt0]/max(me$death))^3
me$weightpoly4[wt0] <- (me$death[wt0]/max(me$death))^4
me$weightpoly5[wt0] <- (me$death[wt0]/max(me$death))^5
me$weightpoly6[wt0] <- (me$death[wt0]/max(me$death))^6
me$dosegroup <- me$group
levels(me$dosegroup) <- c("0", "37", "75", "150")
me$dose <- as.numeric(as.character(me$dosegroup))
# Notice, the modell use an identity link for RD,
fitpoly1 <- glm(tumour ~ dose, data=me, family=binomial(link="identity"), weight=weightpoly1)
fitpoly2 <- glm(tumour ~ dose, data=me, family=binomial(link="identity"), weight=weightpoly2)
fitpoly3 <- glm(tumour ~ dose, data=me, family=binomial(link="identity"), weight=weightpoly3)
fitpoly4 <- glm(tumour ~ dose, data=me, family=binomial(link="identity"), weight=weightpoly4)
fitpoly5 <- glm(tumour ~ dose, data=me, family=binomial(link="identity"), weight=weightpoly5)
fitpoly6 <- glm(tumour ~ dose, data=me, family=binomial(link="identity"), weight=weightpoly6)
library(tukeytrend)
library(multcomp)
ttpoly1 <- tukeytrendfit(fitpoly1, dose="dose", scaling="ari")
ttpoly2 <- tukeytrendfit(fitpoly2, dose="dose", scaling="ari")
ttpoly3 <- tukeytrendfit(fitpoly3, dose="dose", scaling="ari")
ttpoly4 <- tukeytrendfit(fitpoly4, dose="dose", scaling="ari")
ttpoly5 <- tukeytrendfit(fitpoly5, dose="dose", scaling="ari")
ttpoly6 <- tukeytrendfit(fitpoly6, dose="dose", scaling="ari")
compttpoly1 <- glht(model=ttpoly1$mmm, linfct=ttpoly3$mlf, alternative="greater")
compttpoly2 <- glht(model=ttpoly2$mmm, linfct=ttpoly3$mlf, alternative="greater")
compttpoly3 <- glht(model=ttpoly3$mmm, linfct=ttpoly3$mlf, alternative="greater")
compttpoly4 <- glht(model=ttpoly4$mmm, linfct=ttpoly3$mlf, alternative="greater")
compttpoly5 <- glht(model=ttpoly5$mmm, linfct=ttpoly3$mlf, alternative="greater")
compttpoly6 <- glht(model=ttpoly6$mmm, linfct=ttpoly3$mlf, alternative="greater")
tt16 <- combtt(ttpoly1, ttpoly2, ttpoly3, ttpoly4, ttpoly5, ttpoly6)
TT16 <- summary(asglht(tt16, alternative="greater"))
```

5.2 R-code: Tukey-Williams poly-k=5 trend tests

```
#####
# k=5 Tukey-Williams test
TW5 <- tukeytrendfit(fitpoly5, dose="dose",
                      scaling=c("ari", "ord", "arilog", "treat"), ctype="Williams")
tw5 <- glht(model=TW5$mmm, linfct=TW5$mlf, alternative="greater")
polytW5<-fortify(summary(tw5))$test$pvalue
#### marginal Williams contrast
me$Dose <- as.factor(me$dose)
```

```
fitpoly5a <- glm(tumour ~ Dose, data=me, family=binomial(link="identity"), weight=weightpoly5)
TWi5<-summary(glht(fitpoly5a,linfct= mcp(Dose="Williams"), alternative="greater"), adjusted("none"))
```

5.3 R-code: Williams trend tests for five correlated liver tumors

```
library(multcomp)
glm24 <- glm(cbind(Successes + 1, 52 - (Successes+1)) ~ Group,
subset(miceN, Site=="t24"), family=binomial())
glm27 <- glm(cbind(Successes + 1, 52 - (Successes+1)) ~ Group,
subset(miceN, Site=="t27"), family=binomial())
glm29 <- glm(cbind(Successes + 1, 52 - (Successes+1)) ~ Group,
subset(miceN, Site=="t29"), family=binomial())
glm30 <- glm(cbind(Successes + 1, 52 - (Successes+1)) ~ Group,
subset(miceN, Site=="t30"), family=binomial())
mTW <-summary(glht(mmm(T24=glm24,T27=glm27,T29=glm29,T30=glm30),
mlf(mcp(Group ="Williams")), alternative="greater"))
```

5.4 R-code: Multiple correlated crude tumor incidence using Williams trend tests

```
glm24 <- glm(cbind(Successes + 1, 52 - (Successes+1)) ~ Group,
subset(miceN, Site=="t24"), family=binomial())
glm27 <- glm(cbind(Successes + 1, 52 - (Successes+1)) ~ Group,
subset(miceN, Site=="t27"), family=binomial())
glm29 <- glm(cbind(Successes + 1, 52 - (Successes+1)) ~ Group,
subset(miceN, Site=="t29"), family=binomial())
glm30 <- glm(cbind(Successes + 1, 52 - (Successes+1)) ~ Group,
subset(miceN, Site=="t30"), family=binomial())
mTW <-summary(glht(mmm(T24=glm24,T27=glm27,T29=glm29,T30=glm30),
mlf(mcp(Group ="Williams")), alternative="greater"))
```

5.5 R-code: Multiple effect sizes using poly-k=3 Tukey-type trend tests

```
miceL$weightpoly24 <- 1
wt024 <- which(miceL$t24 == 0)
miceL$weightpoly24[wt024] <- (miceL$death[wt024]/max(miceL$death))^3
library(glm2)
library(logbin, quietly=TRUE)
t24OR <-glm(t24~dose, data=miceL, family= binomial(link="logit"), weight=weightpoly24)
t24RD <-glm2(t24~dose, data=miceL, family= binomial(link="identity"), weight=weightpoly24)
t24RR <-glm2(t24~dose, data=miceL, family= binomial(link="log"), weight=weightpoly24)
tuOR <- tukeytrendfit(t24OR, dose="dose", scaling=c("ari", "ord", "arilog"))
pOR <- summary(glht(model=tuOR$mmm, linfct=tuOR$mlf, alternative="greater"))
tuRD <- tukeytrendfit(t24RD, dose="dose", scaling=c("ari", "ord", "arilog"))
pRD <- summary(glht(model=tuRD$mmm, linfct=tuRD$mlf, alternative="greater"))
tuRR <- tukeytrendfit(t24RR, dose="dose", scaling=c("ari", "ord", "arilog"))
pRR <- summary(glht(model=tuRR$mmm, linfct=tuRR$mlf, alternative="greater"))
ttany <- combtt(tuOR, tuRD,tuRR)
anylink<-summary(glht(model=ttany$mmm, linfct=ttany$mlf, alternative="greater"))
```

5.6 R-code: Pairwise and trend tests jointly: crude tumor incidence of hepatoblastoma

```
#####
# Permutative pairwise and William tests
miceT <-miceF[, c(1, 27), ]# t24 only
miceT$Group <-as.factor(miceT$group)
table(miceT)
library("coin")
Wh <- c(-1, 0, 0, 1)
Wm <- c(-1, 0, .5,.5)
Wp <- c(-1, 1/3, 1/3, 1/3)
p3 <- c(-1,0,0,1)
p2 <- c(-1,0,1,0)
p1 <- c(-1,1,0,0)
g <- function(x) {
  x <- unlist(x)
  cbind( W1 = Wh[x], W2 = Wm[x],W3 = Wp[x], P1=p1[x], P2=p2[x])
}
it <- independence_test(t24~ Group, data = miceT, xtrafo = g, alternative = "greater")
pvalue(it, method = "single-step")
#####
# asymptotic Dunnnett and Tukey
TWDu5 <- tukeytrendfit(fitpoly5, dose="dose",
scaling=c("ari", "ord", "arilog", "treat"), ctype="Dunnett")
twdu5 <- glht(model=TWDu5$mmm, linfct=TWDu5$mlf, alternative="greater")
polytWDu5<-fortify(summary(twdu5))
```

5.7 R-code: Complete example

```

miceZ$weightpoly241 <- 1
miceZ$weightpoly242 <- 1
miceZ$weightpoly243 <- 1
miceZ$weightpoly244 <- 1
miceZ$weightpoly245 <- 1
miceZ$weightpoly246 <- 1
wt024 <- which(miceZ$t24 == 0)
miceZ$weightpoly241[wt024] <- (miceZ$death[wt024]/max(miceZ$death))^1
miceZ$weightpoly242[wt024] <- (miceZ$death[wt024]/max(miceZ$death))^2
miceZ$weightpoly243[wt024] <- (miceZ$death[wt024]/max(miceZ$death))^3
miceZ$weightpoly244[wt024] <- (miceZ$death[wt024]/max(miceZ$death))^4
miceZ$weightpoly245[wt024] <- (miceZ$death[wt024]/max(miceZ$death))^5
miceZ$weightpoly246[wt024] <- (miceZ$death[wt024]/max(miceZ$death))^6
miceZ$weightpoly271 <- 1
miceZ$weightpoly272 <- 1
miceZ$weightpoly273 <- 1
miceZ$weightpoly274 <- 1
miceZ$weightpoly275 <- 1
miceZ$weightpoly276 <- 1
wt027 <- which(miceZ$t27 == 0)
miceZ$weightpoly271[wt027] <- (miceZ$death[wt027]/max(miceZ$death))^1
miceZ$weightpoly272[wt027] <- (miceZ$death[wt027]/max(miceZ$death))^2
miceZ$weightpoly273[wt027] <- (miceZ$death[wt027]/max(miceZ$death))^3
miceZ$weightpoly274[wt027] <- (miceZ$death[wt027]/max(miceZ$death))^4
miceZ$weightpoly275[wt027] <- (miceZ$death[wt027]/max(miceZ$death))^5
miceZ$weightpoly276[wt027] <- (miceZ$death[wt027]/max(miceZ$death))^6
miceZ$weightpoly291 <- 1
miceZ$weightpoly292 <- 1
miceZ$weightpoly293 <- 1
miceZ$weightpoly294 <- 1
miceZ$weightpoly295 <- 1
miceZ$weightpoly296 <- 1
wt029 <- which(miceZ$t29 == 0)
miceZ$weightpoly291[wt029] <- (miceZ$death[wt029]/max(miceZ$death))^1
miceZ$weightpoly292[wt029] <- (miceZ$death[wt029]/max(miceZ$death))^2
miceZ$weightpoly293[wt029] <- (miceZ$death[wt029]/max(miceZ$death))^3
miceZ$weightpoly294[wt029] <- (miceZ$death[wt029]/max(miceZ$death))^4
miceZ$weightpoly295[wt029] <- (miceZ$death[wt029]/max(miceZ$death))^5
miceZ$weightpoly296[wt029] <- (miceZ$death[wt029]/max(miceZ$death))^6

t241OR <-glm(t24~dose, data=miceZ, family= binomial(link="logit"), weight=weightpoly241)
t241RD <-glm(t24~dose, data=miceZ, family= binomial(link="identity"), weight=weightpoly241)
t242OR <-glm(t24~dose, data=miceZ, family= binomial(link="logit"), weight=weightpoly242)
t242RD <-glm(t24~dose, data=miceZ, family= binomial(link="identity"), weight=weightpoly242)
t243OR <-glm(t24~dose, data=miceZ, family= binomial(link="logit"), weight=weightpoly243)
t243RD <-glm(t24~dose, data=miceZ, family= binomial(link="identity"), weight=weightpoly243)
t244OR <-glm(t24~dose, data=miceZ, family= binomial(link="logit"), weight=weightpoly244)
t244RD <-glm(t24~dose, data=miceZ, family= binomial(link="identity"), weight=weightpoly244)
t245OR <-glm(t24~dose, data=miceZ, family= binomial(link="logit"), weight=weightpoly245)
t245RD <-glm(t24~dose, data=miceZ, family= binomial(link="identity"), weight=weightpoly245)
t246OR <-glm(t24~dose, data=miceZ, family= binomial(link="logit"), weight=weightpoly246)
t246RD <-glm(t24~dose, data=miceZ, family= binomial(link="identity"), weight=weightpoly246)
t271OR <-glm(t27~dose, data=miceZ, family= binomial(link="logit"), weight=weightpoly271)
t272OR <-glm(t27~dose, data=miceZ, family= binomial(link="logit"), weight=weightpoly272)
t273OR <-glm(t27~dose, data=miceZ, family= binomial(link="logit"), weight=weightpoly273)
t274OR <-glm(t27~dose, data=miceZ, family= binomial(link="logit"), weight=weightpoly274)
t275OR <-glm(t27~dose, data=miceZ, family= binomial(link="logit"), weight=weightpoly275)
t276OR <-glm(t27~dose, data=miceZ, family= binomial(link="logit"), weight=weightpoly276)
t271RD <-glm(t27~dose, data=miceZ, family= binomial(link="identity"), weight=weightpoly271)
t272RD <-glm(t27~dose, data=miceZ, family= binomial(link="identity"), weight=weightpoly272)
t273RD <-glm(t27~dose, data=miceZ, family= binomial(link="identity"), weight=weightpoly273)
t274RD <-glm(t27~dose, data=miceZ, family= binomial(link="identity"), weight=weightpoly274)
t275RD <-glm(t27~dose, data=miceZ, family= binomial(link="identity"), weight=weightpoly275)
t276RD <-glm(t27~dose, data=miceZ, family= binomial(link="identity"), weight=weightpoly276)
t291OR <-glm(t29~dose, data=miceZ, family= binomial(link="logit"), weight=weightpoly291)
t292OR <-glm(t29~dose, data=miceZ, family= binomial(link="logit"), weight=weightpoly292)
t293OR <-glm(t29~dose, data=miceZ, family= binomial(link="logit"), weight=weightpoly293)
t294OR <-glm(t29~dose, data=miceZ, family= binomial(link="logit"), weight=weightpoly294)
t295OR <-glm(t29~dose, data=miceZ, family= binomial(link="logit"), weight=weightpoly295)
t296OR <-glm(t29~dose, data=miceZ, family= binomial(link="logit"), weight=weightpoly296)
t291RD <-glm(t29~dose, data=miceZ, family= binomial(link="identity"), weight=weightpoly291)
t292RD <-glm(t29~dose, data=miceZ, family= binomial(link="identity"), weight=weightpoly292)
t293RD <-glm(t29~dose, data=miceZ, family= binomial(link="identity"), weight=weightpoly293)
t294RD <-glm(t29~dose, data=miceZ, family= binomial(link="identity"), weight=weightpoly294)
t295RD <-glm(t29~dose, data=miceZ, family= binomial(link="identity"), weight=weightpoly295)
t296RD <-glm(t29~dose, data=miceZ, family= binomial(link="identity"), weight=weightpoly296)

library(tukeytrend)
tuOR241 <- tukeytrendfit(t241OR, dose="dose", scaling=c("ari", "ord", "arilog"))
tuOR242 <- tukeytrendfit(t242OR, dose="dose", scaling=c("ari", "ord", "arilog"))
tuOR243 <- tukeytrendfit(t243OR, dose="dose", scaling=c("ari", "ord", "arilog"))
tuOR244 <- tukeytrendfit(t244OR, dose="dose", scaling=c("ari", "ord", "arilog"))
tuOR245 <- tukeytrendfit(t245OR, dose="dose", scaling=c("ari", "ord", "arilog"))
tuOR246 <- tukeytrendfit(t246OR, dose="dose", scaling=c("ari", "ord", "arilog"))
tuRD241 <- tukeytrendfit(t241RD, dose="dose", scaling=c("ari", "ord", "arilog"))
tuRD242 <- tukeytrendfit(t242RD, dose="dose", scaling=c("ari", "ord", "arilog"))
tuRD243 <- tukeytrendfit(t243RD, dose="dose", scaling=c("ari", "ord", "arilog"))
tuRD244 <- tukeytrendfit(t244RD, dose="dose", scaling=c("ari", "ord", "arilog"))
tuRD245 <- tukeytrendfit(t245RD, dose="dose", scaling=c("ari", "ord", "arilog"))
tuRD246 <- tukeytrendfit(t246RD, dose="dose", scaling=c("ari", "ord", "arilog"))
tuOR271 <- tukeytrendfit(t271OR, dose="dose", scaling=c("ari", "ord", "arilog"))
tuOR272 <- tukeytrendfit(t272OR, dose="dose", scaling=c("ari", "ord", "arilog"))
tuOR273 <- tukeytrendfit(t273OR, dose="dose", scaling=c("ari", "ord", "arilog"))
tuOR274 <- tukeytrendfit(t274OR, dose="dose", scaling=c("ari", "ord", "arilog"))
tuRD275 <- tukeytrendfit(t275RD, dose="dose", scaling=c("ari", "ord", "arilog"))
tuRD276 <- tukeytrendfit(t276RD, dose="dose", scaling=c("ari", "ord", "arilog"))
tuOR291 <- tukeytrendfit(t291OR, dose="dose", scaling=c("ari", "ord", "arilog"))
tuOR292 <- tukeytrendfit(t292OR, dose="dose", scaling=c("ari", "ord", "arilog"))
tuOR293 <- tukeytrendfit(t293OR, dose="dose", scaling=c("ari", "ord", "arilog"))
tuOR294 <- tukeytrendfit(t294OR, dose="dose", scaling=c("ari", "ord", "arilog"))

```

```

tuOR295 <- tukeytrendfit(t295OR, dose="dose", scaling=c("ari", "ord", "arilog"))
tuOR296 <- tukeytrendfit(t296OR, dose="dose", scaling=c("ari", "ord", "arilog"))
tuRD291 <- tukeytrendfit(t291RD, dose="dose", scaling=c("ari", "ord", "arilog"))
tuRD292 <- tukeytrendfit(t292RD, dose="dose", scaling=c("ari", "ord", "arilog"))
tuRD293 <- tukeytrendfit(t293RD, dose="dose", scaling=c("ari", "ord", "arilog"))
tuRD294 <- tukeytrendfit(t294RD, dose="dose", scaling=c("ari", "ord", "arilog"))
tuRD295 <- tukeytrendfit(t295RD, dose="dose", scaling=c("ari", "ord", "arilog"))
tuRD296 <- tukeytrendfit(t296RD, dose="dose", scaling=c("ari", "ord", "arilog"))

ttall <- combtt(tuOR241, tuOR242, tuOR243, tuOR244, tuOR245, tuOR246,
                 tuRD241, tuRD242, tuRD243, tuRD244, tuRD245, tuRD246,
                 tuOR271, tuOR272, tuOR273, tuOR274, tuOR275, tuOR276,
                 tuRD271, tuRD272, tuRD273, tuRD274, tuRD275, tuRD276,
                 tuRD291, tuRD292, tuRD293, tuRD294, tuRD295, tuRD296,
                 tuOR291, tuOR292, tuOR293, tuOR294, tuOR295, tuOR296)

all<-summary(asglht(ttall))
allU<-summary(asglht(ttall), adjusted="none")
library(ggplot2)
ALLU<-fortify(allU)
ALL<-fortify(all)
GEM<-cbind(ALL, ALLU$p)

```

5.8 R-code: Complete example

```

ttrD <- combtt(tuRD241, tuRD242, tuRD243, tuRD244, tuRD245, tuRD246,
                 tuRD271, tuRD272, tuRD273, tuRD274, tuRD275, tuRD276,
                 tuRD291, tuRD292, tuRD293, tuRD294, tuRD295, tuRD296)
TTRD<-asglht(ttrD)
plot(TTRD, xlab="Risk difference to p_0", abline=0, main="Simultaneous confidence intervals")

```

5.9 R-code: Exact tests

```

miceZ <-miceF[, c(1:3, 22:38,90)]
miceZ$Group<-as.factor(miceZ$group)
##### Dunnett conditional
library("coin")
aa <- c(-1,0,0,1)
dd <- c(-1,0,1,0)
rr <- c(-1,1,0,0)
g <- function(x) {
  x <- unlist(x)
  cbind( CD1 = rr[x], CD2 = dd[x],CD3 = aa[x])
}
it <- independence_test(t24+t27+t29~ Group, data = miceZ,
                         xtrafo = g, alternative = "greater")
pvalM <- round(pvalue(it, method = "single-step"), 14)
##### add1 asymptotic
library(multcomp)
library(arm)
glm24 <- bayesglm(cbind(t24 + .5, 52 - (t27+.5)) ~ Group,data=miceZ,
                    family=binomial("identity"))
glm27 <- bayesglm(cbind(t27 + .5, 52 - (t27+.5)) ~ Group,data=miceZ,
                    family=binomial("identity"))
glm29 <- bayesglm(cbind(t29 + .5, 52 - (t29+.5)) ~ Group,data=miceZ,
                    family=binomial("identity"))
multT <-summary(glht(mmm(T24=glm24, T27=glm27,T29=glm29),
                      mlf(mcp(Group ="Dunnett"))), alternative="greater")
#####multfisher
library(multfisher)
##### pairwise data 0 vs. 1
miceX <-miceF[, c(1,27,30,32)] # crude proportions
MICE03<-droplevels(subset(miceX,group %in% c(0,3)))
MICE03$group[MICE03$group==3]<-1
MICE02<-subset(miceX,group %in% c(0,2))
MICE02$group[MICE02$group==2]<-1
MICE01<-subset(miceX,group %in% c(0,1))

mf01<-mfisher.test(x=MICE01[,c(2,3,4)],y=MICE01$group,method="alpha",
                     closed.test=TRUE,show.region=TRUE, alpha=0.05)
mf02<-mfisher.test(x=MICE02[,c(2,3,4)],y=MICE02$group,method="number",
                     closed.test=TRUE,show.region=TRUE, alpha=0.05)
mf03<-mfisher.test(x=MICE03[,c(2,3,4)],y=MICE03$group,method="alpha.greedy",
                     closed.test=TRUE,show.region=TRUE, alpha=0.05)
round(mf01$elementary.tests$p.adj,digits=19)
round(mf01$closed.test$p.intersect,digits=19)

```