Leibniz
Universität Hannover

Reports of the Institute of Biostatistics

No 02 / 2007

Leibniz University of Hannover

Natural Sciences Faculty

Approximate simultaneous confidence intervals for the evaluation of long term carcinogenicity trials

Authors: Sill, M. Schaarschmidt, F.

0.1 Notations

Assume that animals are randomly assigned to I treatment groups. In the begin of the study, each treatment group contains n_i animals, where $i = 2, \cdots, I$ are the dose groups, and $i = 1$ is the control group. Let t_{ij} denote the time of death of the jth animal in the ith group, where $j = 1, \cdots, n_i$, and let t_{max} denote the maximal time of death among all animals, usually the time of the final sacrifice at the end of the observation period. Further y_{ij} is the tumor status of the jth animal in the ith treatment group at time of death, with $y_{ij} = 0$ if no tumor is present and $y_{ij} = 1$ otherwise.

0.2 Survival adjustment

Further, the observed proportion of tumor bearing animals over the whole period of observation $p_i = y_{i.}/n_i$, with $y_{i.} = \sum_{j=1}^{n_i} y_{ij}$, is a crude estimator measuring the carcinogenic effect of the compound. This estimator does not take into account, that the time under risk might differ systematically depending on the treatment level. To account for censoring due to treatment-specific mortality, Bailer and Portier (1988) proposed the poly-3 adjustment. Let T denote a continuous random variable, representing the time to the events tumor onset $Y(t) = 1$ or death without tumor $Y(t) = 0$. Under this assumption the tumor incidence rate and the treatment mortality rate can be expressed as event specific hazard functions (Peddada et al., 2005), where

$$\lambda_i(t) = \lim_{\epsilon \to 0} \{P(t \leq T < t + \epsilon, Y(T) = 1 | T \geq t)/\epsilon)\}$$

is the tumor incidence rate and

$$\beta_i(t) = \lim_{\epsilon \to 0} \{P(t \leq T < t + \epsilon, Y(T) = 0 | T \geq t)/\epsilon)\}$$

is the mortality rate.

The observable parameter is the proportion of animals that develop a tumor during the study, which can be expressed as a function of the two hazard rates.

$$p_i = \int_{t=0}^{t=t_{max}} \lambda_i(t) \exp\left[- \int_{s=0}^{s=t} (\lambda_i(s) + \beta_i(s)) \, \mathrm{d}s \right] \mathrm{d}t$$

To adjust for the confounding factor of different mortality over the treatment groups Bailer and Portier (1988) introduced individual weights w_{ij}, taking the different times under risk into account. The weight takes the value $w_{ij} = 1$, if an animal survives until the final sacrifice ($t_{ij} = t_{max}$) or dies during the study with a tumor present. In cases where animals die prior to the terminal sacrifice and no tumors are present, the weight has the value $w_{ij} = (t_{ij}/t_{max})^k$. Bailer and Portier fixed the exponent of the weight k to 3, because it was found that tumors often occur at the rate of a third- to fifth-order polynomial in time (Portier et al., 1986). Summing up, the weights result in an adjusted group sample size $n_i^* = \sum_{j=1}^{n_i} w_{ij}$. An adjusted estimator for the tumor rate per group then is $p_i^* = y_i/n_i^*$, which can be interpreted as estimator for the uncensored tumor incidence rate λ_i. Note, that the validity of the following methods depends on the appropriateness of the poly-k adjustment.

0.3 Global hypotheses for trend tests

We are interested in testing the global null hypothesis of equal tumor incidence rates over the I treatment groups:

$$H_0 : \lambda_1 = \cdots = \lambda_I \tag{1}$$

against the one-sided alternative of increasing tumor rates:

$$H_1 : \lambda_1 \leq \cdots \leq \lambda_I \tag{2}$$

We consider one-sided hypotheses and therefore one-sided tests and one-sided confidence limits only, because only increasing tumor rates is of interest. Two-sided procedures result in higher false negative rates, which should avoided in toxicological risk assessment.
Bretz (2006) showed, how the Williams test for trend among I ordered means $\mu_1, \mu_2, ..., \mu_I$ of Gaussian random variables can be decomposed to a multiple contrast test of $M = I - 1$ contrasts in the general unbalanced case. Hothorn and Bretz (2000) and Bretz and Hothorn (2003) demonstrated the availability of other trend tests if hypotheses are expressed in terms of multiple contrasts, and extended these methods to binomial data (Bretz and Hothorn, 2002). The hypotheses in (1) and (2) can be decomposed to tests of M linear contrasts $L_m = \sum_{i=1}^{I} c_{mi}\lambda_i$. The coefficients c_{mi} of all single contrasts fulfill the condition $\sum_{i=1}^{I} c_{mi} = 0$. Moreover, we will choose c_{mi} such that the condition $\sum_{i:c_{mi}\leq 1} |c_{mi}| = \sum_{i:c_{mi}\geq 1} c_{mi} = 1$ is fulfilled. Then, the contrast L_m can be interpreted as difference of weighted averages of λ_i. The global null hypothesis then is expressed as $\bigcap_{m=1}^{M} L_m = 0$, while the alternative is $\bigcup_{m=1}^{M} L_m > 0$. As an example, in Table 1 the $M = 3$ contrasts of a Williams-type contrast test are shown, for $I = 4$ treatments "control", "low", "medium", and "high" and with equal sample sizes n_i.

comparison	contrast coefficients	control	low	medium	high
high vs. control	c_{i1}	-1	0	0	1
high, medium vs. control	c_{i2}	-1	0	0.5	0.5
high, medium, low vs. control	c_{i3}	-1	$0.\bar{3}$	$0.\bar{3}$	$0.\bar{3}$

Table 1: Williams contrasts for a control and 3 dose groups for balanced sample sizes

Using the same framework, other multiple contrasts are available. For the detection of difference between dose levels, a change-point contrast can be used (Hirotsu and Marumo, 2002). Notice, this is a basic contrast, since several multiple contrast tests on trend can be reformulated as the change-point contrast (Hirotsu et al., 2007).

In some long term carcinogenicity studies a downturn effect in the proportion of tumor bearing animals in the higher dose groups can be observed. This downturn effect may be due to increased treatment mortalities in the high dose groups, but it may still be present after the survival-adjustment. In such cases, the monotonicity assumption of order restricted approaches is violated. A multiple contrast which takes such possible downturns into account is the Williams-type downturn-protected contrast (Bretz and Hothorn, 2003).

In situations where the monotonicity assumption is violated, a Dunnett-type contrast (Dunnett, 1955) is a robust alternative to trend tests. Its global null hypothesis is rejected if any of the dose or treatment groups leads to an increased tumor rate. Moreover, it can be of interest to quantify the differences in tumor rates between the dose groups and the control, indicating the use of simultaneous confidence intervals. Moreover, Dunnett procedure belongs to the most used statistical approaches (Ryan and Woodall, 2005) and is widely used for the evaluation of continuous endpoints in repeated toxicity studies. Therefore, a Dunnett-type poly-3 adjusted procedure can be recommended for routine evaluation of long-term carcinogenicity studies as well.

0.4 Approximate simultaneous confidence intervals

The point estimator for the mth linear combination is $\hat{L}_m = \sum_{i=1}^{I} c_{mi} p_i^*$. Assuming normality for large sample sizes, simultaneous lower $(1-\alpha)$-confidence limits for $L_1, \cdots, L_M$ can be constructed as:

$$\sum_{i=1}^{I} c_{mi} p_i^* - z_{M,R,1-\alpha} \sqrt{\sum_{i=1}^{I} c_{mi}^2 \hat{V}\left(p_i^*\right)} \tag{3}$$

where $\hat{V}\left(p_i^*\right)$ is the variance estimator of p_i^*, where different versions are discussed in section 0.5. The value $z_{M,R,1-\alpha}$ is the equicoordinate $(1-\alpha)$-quantile of an M-variate normal distribution with correlation matrix R and CDF $\Phi_M\left(z; \mu, \mathbf{R}\right) = P\left(\mathbf{Z_M} \leq \mathbf{z}; \mathbf{0}, \mathbf{R}\right)$ for all elements in $\mathbf{Z_M}$. $\mathbf{Z_M}$ is a M-variate normal random vector with expectation vector $\mathbf{0}$ of length M, and R is an $M \times M$ correlation matrix. Such critical values can be computed using the function qmvnorm in the R-package mvtnorm (Hothorn et al., 2001)
The correlation between two linear combinations L_m and $L_{m'}$ depends on the known contrast coefficients c_{mi} and $c_{m'i}$, as well as the unknown variance $V\left(p_i^*\right)$. Here, we estimate elements $\rho_{mm'}$ of the correlation matrix R, using the sample estimates for the variance:

$$\hat{\rho}_{mm'} = \frac{\sum_{i=1}^{I} c_{mi} c_{m'i} \hat{V}\left(p_i^*\right)}{\sqrt{\left(\sum_{i=1}^{I} c_{mi}^2 \hat{V}\left(p_i^*\right)\right)\left(\sum_{i=1}^{I} c_{m'i}^2 \hat{V}\left(p_i^*\right)\right)}} \tag{4}$$

Choosing $z_{M,R,1-\alpha}$ results in confidence limits which asymptotically contain all true values $L_1, \cdots, L_M$ with coverage probability $1-\alpha$. The global null hypothesis is rejected with an approximate familywise error rate α, if the lower limit for at least one contrast excludes 0.

0.5 Variance estimators and adjustment for small sample sizes

There are different ways to derive the sample estimates for variance of p_i^* needed in the equations (3) to (4). Bailer and Portier (1988) apply the Cochran-Armitage test for binomial data on the adjusted tumor rates. Following their ideas, we use $\hat{V}\left(p_i^*\right) = p_i^*\left(1-p_i^*\right)/n_i^*$. Confidence intervals using this variance estimator in the equations (3) to (4) will be denoted BP-Wald method.

Bieler and Williams (1993) introduced an improved variance estimator

$$V\left(p_i^*\right) = \left(n_i / \left(n_i - 1\right)\right) \sum_{j=1}^{n_i} \left(d_{ij} - \bar{d}_i\right)^2,$$

with $d_{ij} = \left(y_{ij} - p_i^* w_{ij}\right) / n_{ij}^*$ and $\bar{d}_i = \sum_{j=1}^{n_i} d_{ij} / n_i$. Using this estimator in the equations (3) to (4) leads to methods referred to as BW-Wald in the following. The BP-Wald variance estimator has value 0 in cases that $y_{i\cdot} = 0$ or $y_{i\cdot} = n_i$ is observed and therefore leads to situations where Z_m and the elements of R are not defined and confidence intervals of length 0 result for certain contrasts. For the BW-Wald method this holds true if $y_{i\cdot} = 0$. To use the BP-Wald and BW-Wald estimators in the simulation study, we modified their variance estimator such that $p_i^* = 0.5/n_i^*$ in case that $y_i = 0$ is observed. Further in cases that $y_i = n_i$ is observed we adjusted the BP-Wald estimator by $p_i^* = \left(n_i - 0.5\right) / n_i^*$.

For the difference of binomial proportions, Wald-type confidence intervals have too low coverage probabilities, especially for low proportions and moderate to small sample (Agresti and Caffo, 2000; Price and Bonett, 2004). Since poly-3 adjusted tumor rates share some properties of binomial proportions, we propose to use similar adjustments for small sample sizes. Using

$$p_i^* = \left(y_i + 0.5\right) / \left(n_i^* + 1\right)$$

and

$$\hat{V}\left(p_i^*\right) = p_i^* \left(1 - p_i^*\right) / \left(n_i^* + 1\right)$$

in equations (3) to (4) will be referred to as Add-1 method, using

$$p_i^* = \left(y_i + 1\right) / \left(n_i^* + 2\right)$$

in combination with

$$\hat{V}\left(p_i^*\right) = p_i^* \left(1 - p_i^*\right) / \left(n_i^* + 2\right)$$

will be called Add-2 method. Both adjustment shift the group wise point estimates to 0.5 and increases the group wise variance estimators for small proportions, resulting in wider intervals with a more conservative behavior. The second approach follows the "adding 1 failure and 2 successes" idea of Agresti and Caffo (2000) for the two sample case. Its adaptation for a linear combination of more than two proportion has been investigated by Price and Bonett (2004). Both methods lead to well performing confidence intervals for binomial data, with coverage probability close to the nominal level.

0.6 Simulation study

The simulation study was performed for the commonly used design $n_i = 50$, $I = 4$. Daily records of death and tumor presence were simulated assuming a study period of 730 days. The probability $P\left(T_t \leq t\right)$ to onset an irreversible tumor until time t, and the probability $P\left(T_m \leq t\right)$ to die for any reason until time t, are assumed to follow independent Weibull distributions $P(T_t \leq t) = 1 - \exp(-(t/b_t)^{a_t})$, and $P(T_m \leq t) = 1 - \exp(-(t/b_m)^{a_m})$ using the notation of Johnson et al. (1994). This simplification of the natural process which includes at least three hazard rates, is justified by the facts that no cause of death information is included and an animal dying with presence of a tumor

contributes with weight $w_{ij} = 1$ to the sample estimator, irrespective of its time of death t_{ij} (Peddada et al., 2005). The scale parameters b_t and b_m of the two Weibull distributions were chosen to mimic different mortality patterns at $t = t_{max}$ and different dose-response shapes for the tumor onset. The shape parameters were fixed $a_t = a_m = 3$. As reported by Bailer and Portier (1988), mortality rates of 30% in the control group were considered as usual in long-term gavage experiments.

The Tables shown in this section present estimates for the coverage probability based on 10000 simulation runs for each parameter setting. The nominal confidence level was chosen 95%, the simultaneous coverage probability is defined as the probability that each true value L_m falls beyond the corresponding lower bound for all M elements of L. The column mortality rates β_i present the expected proportions of dead animals at time of the final sacrifice $P(t = 730)$ in the four treatment groups, the column tumor rate displays the parameter λ_i in the four treatment groups $i = 1, \cdots, 4$. In the following, we will display the results for the BP-Wald, BW-Wald, Add-1 and Add-2 methods using the Williams contrast (Table 2) in situations with equal and increasing tumor incidence rates and different mortality patterns. Additionally, we present similar results for the change-point (Table 3) , the Dunnett-type contrast (Table 4) and the Williams-type downturn-protected contrast (Table 5).

mortality rates β_i	tumor rates λ_i	BP-Wald	BW-Wald	Add-1	Add-2
0.3, 0.3, 0.3, 0.3	0.05, 0.05, 0.05, 0.05	0.890	0.891	0.950	0.971
0.3, 0.4, 0.5, 0.6	0.05, 0.05, 0.05, 0.05	0.894	0.897	0.946	0.975
0.3, 0.5, 0.6, 0.7	0.05, 0.05, 0.05, 0.05	0.902	0.910	0.951	0.973
0.5, 0.4, 0.3, 0.2	0.05, 0.05, 0.05, 0.05	0.849	0.856	0.946	0.981
0.3, 0.3, 0.3, 0.3	0.1, 0.1, 0.1, 0.1	0.928	0.931	0.941	0.957
0.3, 0.4, 0.5, 0.6	0.1, 0.1, 0.1, 0.1	0.933	0.935	0.947	0.957
0.3, 0.5, 0.6, 0.7	0.1, 0.1, 0.1, 0.1	0.935	0.938	0.946	0.957
0.5, 0.4, 0.3, 0.2	0.1, 0.1, 0.1, 0.1	0.920	0.913	0.939	0.953
0.3, 0.3, 0.3, 0.3	0.2, 0.2, 0.2, 0.2	0.935	0.937	0.944	0.953
0.3, 0.4, 0.5, 0.6	0.2, 0.2, 0.2, 0.2	0.950	0.941	0.952	0.953
0.3, 0.5, 0.6, 0.7	0.2, 0.2, 0.2, 0.2	0.945	0.944	0.951	0.957
0.5, 0.4, 0.3, 0.2	0.2, 0.2, 0.2, 0.2	0.931	0.930	0.942	0.949
0.3, 0.3, 0.3, 0.3	0.3, 0.3, 0.3, 0.3	0.944	0.942	0.947	0.949
0.3, 0.4, 0.5, 0.6	0.3, 0.3, 0.3, 0.3	0.948	0.946	0.953	0.951
0.3, 0.5, 0.6, 0.7	0.3, 0.3, 0.3, 0.3	0.953	0.965	0.952	0.957
0.5, 0.4, 0.3, 0.2	0.3, 0.3, 0.3, 0.3	0.939	0.938	0.946	0.950
0.3, 0.3, 0.3, 0.3	0.05, 0.1, 0.15, 0.2	0.942	0.941	0.963	0.976
0.3, 0.4, 0.5, 0.6	0.05, 0.1, 0.15, 0.2	0.947	0.946	0.967	0.974
0.3, 0.5, 0.6, 0.7	0.05, 0.1, 0.15, 0.2	0.949	0.953	0.970	0.976
0.6, 0.5, 0.4, 0.3	0.05, 0.1, 0.15, 0.2	0.929	0.929	0.962	0.984
0.3, 0.3, 0.3, 0.3	0.1, 0.25, 0.3, 0.3	0.940	0.941	0.954	0.967
0.3, 0.4, 0.5, 0.6	0.1, 0.25, 0.3, 0.3	0.948	0.946	0.962	0.975
0.3, 0.5, 0.6, 0.7	0.1, 0.25, 0.3, 0.3	0.958	0.950	0.964	0.972
0.6, 0.5, 0.4, 0.3	0.1, 0.25, 0.3, 0.3	0.930	0.931	0.952	0.972
0.3, 0.3, 0.3, 0.3	0.2, 0.2, 0.25, 0.4	0.937	0.939	0.949	0.955
0.3, 0.4, 0.5, 0.6	0.2, 0.2, 0.25, 0.4	0.948	0.945	0.958	0.964
0.3, 0.5, 0.6, 0.7	0.2, 0.2, 0.25, 0.4	0.954	0.951	0.961	0.968

mortality rates β_i	tumor rates λ_i	BP-Wald	BW-Wald	Add-1	Add-2
0.6, 0.5, 0.4, 0.3	0.2, 0.2, 0.25, 0.4	0.936	0.932	0.945	0.957
0.3, 0.3, 0.3, 0.3	0.1, 0.2, 0.3, 0.4	0.940	0.936	0.959	0.970
0.3, 0.4, 0.5, 0.6	0.1, 0.2, 0.3, 0.4	0.950	0.951	0.965	0.973
0.3, 0.5, 0.6, 0.7	0.1, 0.2, 0.3, 0.4	0.958	0.954	0.970	0.977
0.6, 0.5, 0.4, 0.3	0.1, 0.2, 0.3, 0.4	0.939	0.933	0.955	0.969

Table 2: Coverage probabilities of lower 95% confidence intervals for the Williams contrast in settings with equal and increasing tumor rates and different mortality patterns

mortality rates β_i	tumor rates λ_i	BP-Wald	BW-Wald	Add-1	Add-2
0.3, 0.3, 0.3, 0.3	0.05, 0.05, 0.05, 0.05	0.899	0.907	0.955	0.981
0.3, 0.4, 0.5, 0.6	0.05, 0.05, 0.05, 0.05	0.908	0.902	0.957	0.977
0.3, 0.5, 0.6, 0.7	0.05, 0.05, 0.05, 0.05	0.915	0.912	0.960	0.981
0.5, 0.4, 0.3, 0.2	0.05, 0.05, 0.05, 0.05	0.864	0.858	0.960	0.989
0.3, 0.3, 0.3, 0.3	0.1, 0.1, 0.1, 0.1	0.925	0.932	0.944	0.959
0.3, 0.4, 0.5, 0.6	0.1, 0.1, 0.1, 0.1	0.939	0.937	0.947	0.962
0.3, 0.5, 0.6, 0.7	0.1, 0.1, 0.1, 0.1	0.940	0.945	0.952	0.965
0.5, 0.4, 0.3, 0.2	0.1, 0.1, 0.1, 0.1	0.918	0.920	0.944	0.967
0.3, 0.3, 0.3, 0.3	0.2, 0.2, 0.2, 0.2	0.940	0.937	0.949	0.952
0.3, 0.4, 0.5, 0.6	0.2, 0.2, 0.2, 0.2	0.945	0.947	0.953	0.956
0.3, 0.5, 0.6, 0.7	0.2, 0.2, 0.2, 0.2	0.946	0.946	0.956	0.957
0.5, 0.4, 0.3, 0.2	0.2, 0.2, 0.2, 0.2	0.937	0.935	0.947	0.953
0.3, 0.3, 0.3, 0.3	0.3, 0.3, 0.3, 0.3	0.942	0.943	0.948	0.951
0.3, 0.4, 0.5, 0.6	0.3, 0.3, 0.3, 0.3	0.947	0.949	0.958	0.957
0.3, 0.5, 0.6, 0.7	0.3, 0.3, 0.3, 0.3	0.951	0.962	0.960	0.958
0.5, 0.4, 0.3, 0.2	0.3, 0.3, 0.3, 0.3	0.938	0.938	0.946	0.951
0.3, 0.3, 0.3, 0.3	0.05, 0.1, 0.15, 0.2	0.942	0.943	0.962	0.977
0.3, 0.4, 0.5, 0.6	0.05, 0.1, 0.15, 0.2	0.950	0.951	0.965	0.978
0.6, 0.5, 0.4, 0.3	0.05, 0.1, 0.15, 0.2	0.930	0.926	0.968	0.983
0.6, 0.5, 0.4, 0.3	0.05, 0.1, 0.15, 0.2	0.907	0.911	0.955	0.979
0.3, 0.3, 0.3, 0.3	0.1, 0.25, 0.3, 0.3	0.940	0.942	0.955	0.967
0.3, 0.4, 0.5, 0.6	0.1, 0.25, 0.3, 0.3	0.952	0.946	0.962	0.970
0.3, 0.5, 0.6, 0.7	0.1, 0.25, 0.3, 0.3	0.954	0.950	0.966	0.975
0.6, 0.5, 0.4, 0.3	0.1, 0.25, 0.3, 0.3	0.931	0.935	0.953	0.969
0.3, 0.3, 0.3, 0.3	0.2, 0.2, 0.25, 0.4	0.941	0.940	0.948	0.960
0.3, 0.4, 0.5, 0.6	0.2, 0.2, 0.25, 0.4	0.954	0.946	0.957	0.967
0.3, 0.5, 0.6, 0.7	0.2, 0.2, 0.25, 0.4	0.955	0.954	0.966	0.969
0.6, 0.5, 0.4, 0.3	0.2, 0.2, 0.25, 0.4	0.936	0.935	0.949	0.960
0.3, 0.3, 0.3, 0.3	0.1, 0.2, 0.3, 0.4	0.940	0.938	0.958	0.968
0.3, 0.4, 0.5, 0.6	0.1, 0.2, 0.3, 0.4	0.951	0.951	0.964	0.975
0.3, 0.5, 0.6, 0.7	0.1, 0.2, 0.3, 0.4	0.957	0.953	0.968	0.979

mortality rates β_i	tumor rates λ_i	BP-Wald	BW-Wald	Add-1	Add-2
0.6, 0.5, 0.4, 0.3	0.1, 0.2, 0.3, 0.4	0.934	0.931	0.958	0.972

Table 3: Coverage probabilities of lower 95% confidence intervals for the Changepoint contrast in settings with equal and increasing tumor rates and different mortality patterns

mortality rates β_i	tumor rates λ_i	BP-Wald	BW-Wald	Add-1	Add-2
0.3, 0.3, 0.3, 0.3	0.05, 0.05, 0.05, 0.05	0.972	0.975	0.980	0.992
0.3, 0.4, 0.5, 0.6	0.05, 0.05, 0.05, 0.05	0.970	0.979	0.985	0.990
0.3, 0.5, 0.6, 0.7	0.05, 0.05, 0.05, 0.05	0.970	0.976	0.988	0.992
0.5, 0.4, 0.3, 0.2	0.05, 0.05, 0.05, 0.05	0.954	0.971	0.982	0.993
0.3, 0.3, 0.3, 0.3	0.1, 0.1, 0.1, 0.1	0.954	0.959	0.963	0.972
0.3, 0.4, 0.5, 0.6	0.1, 0.1, 0.1, 0.1	0.955	0.960	0.968	0.975
0.3, 0.5, 0.6, 0.7	0.1, 0.1, 0.1, 0.1	0.963	0.962	0.973	0.980
0.5, 0.4, 0.3, 0.2	0.1, 0.1, 0.1, 0.1	0.944	0.944	0.962	0.977
0.3, 0.3, 0.3, 0.3	0.2, 0.2, 0.2, 0.2	0.948	0.949	0.954	0.957
0.3, 0.4, 0.5, 0.6	0.2, 0.2, 0.2, 0.2	0.953	0.953	0.959	0.962
0.3, 0.5, 0.6, 0.7	0.2, 0.2, 0.2, 0.2	0.959	0.954	0.962	0.968
0.5, 0.4, 0.3, 0.2	0.2, 0.2, 0.2, 0.2	0.944	0.943	0.953	0.961
0.3, 0.3, 0.3, 0.3	0.3, 0.3, 0.3, 0.3	0.946	0.949	0.954	0.951
0.3, 0.4, 0.5, 0.6	0.3, 0.3, 0.3, 0.3	0.954	0.950	0.959	0.956
0.3, 0.5, 0.6, 0.7	0.3, 0.3, 0.3, 0.3	0.958	0.966	0.957	0.963
0.5, 0.4, 0.3, 0.2	0.3, 0.3, 0.3, 0.3	0.943	0.945	0.952	0.955
0.3, 0.3, 0.3, 0.3	0.05, 0.1, 0.15, 0.2	0.969	0.970	0.976	0.984
0.3, 0.4, 0.5, 0.6	0.05, 0.1, 0.15, 0.2	0.974	0.971	0.978	0.984
0.3, 0.5, 0.6, 0.7	0.05, 0.1, 0.15, 0.2	0.973	0.972	0.982	0.988
0.6, 0.5, 0.4, 0.3	0.05, 0.1, 0.15, 0.2	0.963	0.961	0.981	0.988
0.3, 0.3, 0.3, 0.3	0.1, 0.25, 0.3, 0.3	0.951	0.950	0.965	0.973
0.3, 0.4, 0.5, 0.6	0.1, 0.25, 0.3, 0.3	0.960	0.953	0.971	0.976
0.3, 0.5, 0.6, 0.7	0.1, 0.25, 0.3, 0.3	0.962	0.955	0.970	0.976
0.6, 0.5, 0.4, 0.3	0.1, 0.25, 0.3, 0.3	0.952	0.947	0.964	0.977
0.3, 0.3, 0.3, 0.3	0.2, 0.2, 0.25, 0.4	0.947	0.946	0.955	0.963
0.3, 0.4, 0.5, 0.6	0.2, 0.2, 0.25, 0.4	0.954	0.953	0.962	0.965
0.3, 0.5, 0.6, 0.7	0.2, 0.2, 0.25, 0.4	0.961	0.952	0.965	0.972
0.6, 0.5, 0.4, 0.3	0.2, 0.2, 0.25, 0.4	0.942	0.941	0.955	0.964
0.3, 0.3, 0.3, 0.3	0.1, 0.2, 0.3, 0.4	0.953	0.952	0.965	0.972
0.3, 0.4, 0.5, 0.6	0.1, 0.2, 0.3, 0.4	0.960	0.953	0.971	0.975
0.3, 0.5, 0.6, 0.7	0.1, 0.2, 0.3, 0.4	0.963	0.958	0.974	0.978
0.6, 0.5, 0.4, 0.3	0.1, 0.2, 0.3, 0.4	0.948	0.946	0.965	0.977

mortality rates β_i	tumor rates λ_i	BP-Wald	BW-Wald	Add-1	Add-2

Table 4: Coverage probabilities of lower 95% confidence intervals for the Dunnett contrast in settings with equal and increasing tumor rates and different mortality patterns

mortality rates β_i	tumor rates λ_i	BP-Wald	BW-Wald	Add-1	Add-2
0.3, 0.3, 0.3, 0.3	0.05, 0.05, 0.05, 0.05	0.862	0.866	0.962	0.988
0.3, 0.4, 0.5, 0.6	0.05, 0.05, 0.05, 0.05	0.860	0.854	0.959	0.988
0.3, 0.5, 0.6, 0.7	0.05, 0.05, 0.05, 0.05	0.859	0.861	0.964	0.986
0.5, 0.4, 0.3, 0.2	0.05, 0.05, 0.05, 0.05	0.825	0.823	0.960	0.993
0.3, 0.3, 0.3, 0.3	0.1, 0.1, 0.1, 0.1	0.908	0.913	0.946	0.969
0.3, 0.4, 0.5, 0.6	0.1, 0.1, 0.1, 0.1	0.918	0.917	0.951	0.968
0.3, 0.5, 0.6, 0.7	0.1, 0.1, 0.1, 0.1	0.923	0.917	0.952	0.967
0.5, 0.4, 0.3, 0.2	0.1, 0.1, 0.1, 0.1	0.899	0.905	0.949	0.969
0.3, 0.3, 0.3, 0.3	0.2, 0.2, 0.2, 0.2	0.924	0.927	0.945	0.957
0.3, 0.4, 0.5, 0.6	0.2, 0.2, 0.2, 0.2	0.932	0.932	0.952	0.960
0.3, 0.5, 0.6, 0.7	0.2, 0.2, 0.2, 0.2	0.934	0.933	0.952	0.963
0.5, 0.4, 0.3, 0.2	0.2, 0.2, 0.2, 0.2	0.925	0.918	0.946	0.959
0.3, 0.3, 0.3, 0.3	0.3, 0.3, 0.3, 0.3	0.938	0.935	0.944	0.953
0.3, 0.4, 0.5, 0.6	0.3, 0.3, 0.3, 0.3	0.938	0.936	0.945	0.955
0.3, 0.5, 0.6, 0.7	0.3, 0.3, 0.3, 0.3	0.938	0.943	0.949	0.959
0.5, 0.4, 0.3, 0.2	0.3, 0.3, 0.3, 0.3	0.927	0.931	0.944	0.955
0.3, 0.3, 0.3, 0.3	0.05, 0.1, 0.15, 0.2	0.911	0.923	0.961	0.984
0.3, 0.4, 0.5, 0.6	0.05, 0.1, 0.15, 0.2	0.917	0.921	0.966	0.983
0.3, 0.5, 0.6, 0.7	0.05, 0.01, 0.15, 0.2	0.924	0.924	0.969	0.986
0.6, 0.5, 0.4, 0.3	0.05, 0.01, 0.15, 0.2	0.899	0.896	0.967	0.989
0.3, 0.3, 0.3, 0.3	0.1, 0.25, 0.3, 0.3	0.927	0.925	0.949	0.974
0.3, 0.4, 0.5, 0.6	0.1, 0.25, 0.3, 0.3	0.931	0.928	0.958	0.974
0.3, 0.5, 0.6, 0.7	0.1, 0.25, 0.3, 0.3	0.930	0.933	0.959	0.976
0.6, 0.5, 0.4, 0.3	0.1, 0.25, 0.3, 0.3	0.912	0.912	0.950	0.975
0.3, 0.3, 0.3, 0.3	0.2, 0.2, 0.25, 0.4	0.932	0.925	0.945	0.960
0.3, 0.4, 0.5, 0.6	0.2, 0.2, 0.25, 0.4	0.931	0.929	0.948	0.965
0.3, 0.5, 0.6, 0.7	0.2, 0.2, 0.25, 0.4	0.939	0.936	0.955	0.965
0.6, 0.5, 0.4, 0.3	0.2, 0.2, 0.25, 0.4	0.921	0.921	0.943	0.960
0.3, 0.3, 0.3, 0.3	0.1, 0.2, 0.3, 0.4	0.929	0.925	0.954	0.970
0.3, 0.4, 0.5, 0.6	0.1, 0.2, 0.3, 0.4	0.929	0.930	0.959	0.979
0.3, 0.5, 0.6, 0.7	0.1, 0.2, 0.3, 0.4	0.931	0.928	0.963	0.978
0.6, 0.5, 0.4, 0.3	0.1, 0.2, 0.3, 0.4	0.908	0.909	0.950	0.977

Table 5: Coverage probabilities of lower 95% confidence intervals for the Williams-type downturn-protected contrast in settings with equal and increasing tumor rates and different mortality patterns

The results in Table 2 show that in situations where equal tumor incidence rates exist the Add-1 interval has coverage probability closest to the nominal level. For increasing tumor incidence rates the interval shows a slightly conservative performance. The Add-2 interval usually exceeds the nominal level. Both, the BP-Wald and BW-Wald interval are liberal in cases of equal tumor incidence rates and especially for very small rates. In situations of increasing incidence rates the coverage probability of the Wald intervals approach the nominal level. In most cases, the BW-Wald interval has coverage probability closer to the nominal level, but in general the difference to the BP-Wald interval is negligible. We found similar results for other multiple contrast types except for the Dunnett-type contrast. In this case all proposed methods are conservative, even under the null and especially for small tumor incidence rates.

In order to study the performance of the lower Add-1 confidence limits in situations with smaller sample sizes, a second simulation study was performed. Here, a balanced sample size of 35 for $I = 4$ groups was considered. Table 6 shows the results for different contrast types.

mortality rates β_i	tumor rates λ_i	Dunnett	Williams	Changepoint
0.3, 0.3, 0.3, 0.3	0.05, 0.1, 0.15, 0.2	0.979	0.970	0.971
0.3, 0.4, 0.5, 0.6	0.05, 0.1, 0.15, 0.2	0.982	0.971	0.972
0.3, 0.5, 0.6, 0.7	0.05, 0.1, 0.15, 0.2	0.984	0.974	0.975
0.3, 0.3, 0.3, 0.3	0.1, 0.25, 0.3, 0.3	0.964	0.956	0.955
0.3, 0.4, 0.5, 0.6	0.1, 0.25, 0.3, 0.3	0.968	0.961	0.962
0.3, 0.5, 0.6, 0.7	0.1, 0.25, 0.3, 0.3	0.974	0.965	0.967
0.3, 0.3, 0.3, 0.3	0.2, 0.2, 0.25, 0.4	0.956	0.944	0.949
0.3, 0.4, 0.5, 0.6	0.2, 0.2, 0.25, 0.4	0.963	0.956	0.956
0.3, 0.5, 0.6, 0.7	0.2, 0.2, 0.25, 0.4	0.965	0.960	0.963
0.3, 0.3, 0.3, 0.3	0.1, 0.2, 0.3, 0.4	0.971	0.957	0.950
0.3, 0.4, 0.5, 0.6	0.1, 0.2, 0.3, 0.4	0.970	0.965	0.965
0.3, 0.5, 0.6, 0.7	0.1, 0.2, 0.3, 0.4	0.973	0.972	0.968

Table 6: Coverage probabilities of the lower 95% Add-1 confidence limits under different alternatives and different multiple contrasts

The results for the Williams contrast in Table 6 compared with the results of Table 2 show, that the coverage probability of the Add-1 interval is slightly conservative for small sample sizes. The coverage probabilities of the Add-1 interval for the Williams-type and the change-point contrast are almost similar. In case of the Dunnett-type contrast the coverage probability of the Add-1 interval strongly exceeds the nominal level.
Summarizing, for the typical NTP design with a control and three dose groups and balanced sample sizes of 50, Wald-type approaches should be avoided because of their serious liberality, whereas the simple Add-1 approximation can be recommended for one-sided tests and confidence limits in several multiple contrasts.

0.7 Extension of the Williams contrast by single linear contrast

In order to get a more sensitive performance for linear dose response relationships we recommend an expansion of the Williams contrast by an additional single linear contrast. This corresponds to the considerations of Peddada and Kissling, to combine a linear trend test with a Williams type test. For the multiple contrast in a balanced 4 group design described in equation 1, this additional single contrast has the following form $c_4 = -0.75, -0.25, 0.25, 0.75$. Note that for this multiple contrast the number of single contrasts is $M = I$.

To examine the power to reject the null hypothesis, a second simulation study with the same settings was performed. The power is defined as the probability that at least one interval excludes zero. Table 7 and Table 8 present the results of the simulation study compared with the results for a regular Williams contrast. BP-Wald denotes the Wald type interval with the variance estimator suggested by Bailer and Portier, while BW-Wald is the Wald interval with the variance estimator of Bieler and Williams. Add-1 and Add-2 are the intervals suggested for small sample sizes, both use the variance estimator of Bailer and Portier.

settings		Williams				Williams+linear			
mortality rates β_i	tumor rates λ_i	BP-Wald	BW-Wald	Add-1	Add-2	BP-Wald	BW-Wald	Add-1	Add-2
0.3, 0.3, 0.3, 0.3	0.05, 0.05, 0.05, 0.05	0.890	0.891	0.950	0.971	0.883	0.889	0.947	0.975
0.3, 0.4, 0.5, 0.6	0.05, 0.05, 0.05, 0.05	0.894	0.897	0.946	0.975	0.899	0.905	0.955	0.973
0.3, 0.5, 0.6, 0.7	0.05, 0.05, 0.05, 0.05	0.902	0.910	0.951	0.973	0.904	0.917	0.957	0.972
0.5, 0.4, 0.3, 0.2	0.05, 0.05, 0.05, 0.05	0.849	0.856	0.946	0.981	0.853	0.863	0.949	0.983
0.3, 0.3, 0.3, 0.3	0.1, 0.1, 0.1, 0.1	0.928	0.931	0.941	0.957	0.930	0.925	0.943	0.953
0.3, 0.4, 0.5, 0.6	0.1, 0.1, 0.1, 0.1	0.933	0.935	0.947	0.957	0.936	0.937	0.948	0.961
0.3, 0.5, 0.6, 0.7	0.1, 0.1, 0.1, 0.1	0.935	0.938	0.946	0.957	0.941	0.935	0.953	0.959
0.5, 0.4, 0.3, 0.2	0.1, 0.1, 0.1, 0.1	0.920	0.913	0.939	0.953	0.919	0.916	0.941	0.959
0.3, 0.3, 0.3, 0.3	0.2, 0.2, 0.2, 0.2	0.935	0.937	0.944	0.953	0.938	0.941	0.948	0.948
0.3, 0.4, 0.5, 0.6	0.2, 0.2, 0.2, 0.2	0.950	0.941	0.952	0.953	0.945	0.940	0.951	0.952
0.3, 0.5, 0.6, 0.7	0.2, 0.2, 0.2, 0.2	0.945	0.944	0.951	0.957	0.951	0.945	0.950	0.953
0.5, 0.4, 0.3, 0.2	0.2, 0.2, 0.2, 0.2	0.931	0.930	0.942	0.949	0.933	0.930	0.940	0.952
0.3, 0.3, 0.3, 0.3	0.3, 0.3, 0.3, 0.3	0.944	0.942	0.947	0.949	0.940	0.941	0.944	0.951
0.3, 0.4, 0.5, 0.6	0.3, 0.3, 0.3, 0.3	0.948	0.946	0.953	0.951	0.952	0.947	0.952	0.954
0.3, 0.5, 0.6, 0.7	0.3, 0.3, 0.3, 0.3	0.953	0.965	0.952	0.957	0.951	0.968	0.957	0.960
0.5, 0.4, 0.3, 0.2	0.3, 0.3, 0.3, 0.3	0.939	0.938	0.946	0.950	0.938	0.936	0.940	0.946
0.3, 0.3, 0.3, 0.3	0.05, 0.1, 0.15, 0.2	0.942	0.941	0.963	0.976	0.935	0.945	0.959	0.975
0.3, 0.4, 0.5, 0.6	0.05, 0.1, 0.15, 0.2	0.947	0.946	0.967	0.974	0.951	0.949	0.964	0.978
0.3, 0.5, 0.6, 0.7	0.05, 0.1, 0.15, 0.2	0.949	0.953	0.970	0.976	0.955	0.951	0.966	0.979
0.6, 0.5, 0.4, 0.3	0.05, 0.1, 0.15, 0.2	0.929	0.929	0.962	0.984	0.928	0.933	0.966	0.984
0.3, 0.3, 0.3, 0.3	0.1, 0.25, 0.3, 0.3	0.940	0.941	0.954	0.967	0.939	0.940	0.955	0.968
0.3, 0.4, 0.5, 0.6	0.1, 0.25, 0.3, 0.3	0.948	0.946	0.962	0.975	0.948	0.942	0.964	0.973
0.3, 0.5, 0.6, 0.7	0.1, 0.25, 0.3, 0.3	0.958	0.950	0.964	0.972	0.955	0.946	0.966	0.973
0.6, 0.5, 0.4, 0.3	0.1, 0.25, 0.3, 0.3	0.930	0.931	0.952	0.972	0.928	0.931	0.954	0.972
0.3, 0.3, 0.3, 0.3	0.2, 0.2, 0.25, 0.4	0.937	0.939	0.949	0.955	0.941	0.940	0.948	0.959
0.3, 0.4, 0.5, 0.6	0.2, 0.2, 0.25, 0.4	0.948	0.945	0.958	0.964	0.950	0.948	0.958	0.964
0.3, 0.5, 0.6, 0.7	0.2, 0.2, 0.25, 0.4	0.954	0.951	0.961	0.968	0.956	0.951	0.962	0.968
0.6, 0.5, 0.4, 0.3	0.2, 0.2, 0.25, 0.4	0.936	0.932	0.945	0.957	0.934	0.933	0.948	0.957
0.3, 0.3, 0.3, 0.3	0.1, 0.2, 0.3, 0.4	0.940	0.936	0.959	0.970	0.941	0.944	0.959	0.970
0.3, 0.4, 0.5, 0.6	0.1, 0.2, 0.3, 0.4	0.950	0.951	0.965	0.973	0.953	0.948	0.965	0.976
0.3, 0.5, 0.6, 0.7	0.1, 0.2, 0.3, 0.4	0.958	0.954	0.970	0.977	0.959	0.956	0.966	0.979
0.6, 0.5, 0.4, 0.3	0.1, 0.2, 0.3, 0.4	0.939	0.933	0.955	0.969	0.932	0.932	0.957	0.975

Table 7: Coverage probabilities of lower 95% confidence intervals of the Williams contrast and the Willimas contrast extended by a single linear contrast in settings with equal and increasing tumor rates and different mortality patterns

The presented results show that in many cases the Wald type intervals have coverage probability below the nominal level. Further the difference between the different variance estimators used for the Wald intervals is in most cases negligible. The coverage probability of the Add-2 interval is too conservative. Therefore we recommend the Add-1 interval which tends to be a little conservative but has most frequently coverage probability nearest to the nominal level. Compared to the regular Williams contrast we found both methods to perform almost equal in terms of coverage probability.

settings		Williams				Williams+linear			
mortality rates β_i	tumor rates λ_i	BP-Wald	BW-Wald	Add-1	Add-2	BP-Wald	BW-Wald	Add-1	Add-2
0.3, 0.3, 0.3, 0.3	0.05, 0.05, 0.05, 0.05	0.121	0.120	0.050	0.029	0.120	0.118	0.049	0.025
0.3, 0.4, 0.5, 0.6	0.05, 0.05, 0.05, 0.05	0.116	0.115	0.051	0.032	0.112	0.116	0.049	0.028
0.3, 0.5, 0.6, 0.7	0.05, 0.05, 0.05, 0.05	0.109	0.109	0.048	0.027	0.105	0.109	0.048	0.026
0.5, 0.4, 0.3, 0.2	0.05, 0.05, 0.05, 0.05	0.156	0.148	0.056	0.021	0.155	0.149	0.050	0.016
0.3, 0.3, 0.3, 0.3	0.1, 0.1, 0.1, 0.1	0.071	0.072	0.056	0.046	0.071	0.071	0.058	0.042
0.3, 0.4, 0.5, 0.6	0.1, 0.1, 0.1, 0.1	0.066	0.073	0.050	0.048	0.068	0.066	0.057	0.041
0.3, 0.5, 0.6, 0.7	0.1, 0.1, 0.1, 0.1	0.065	0.064	0.055	0.047	0.062	0.066	0.048	0.042
0.5, 0.4, 0.3, 0.2	0.1, 0.1, 0.1, 0.1	0.086	0.084	0.059	0.043	0.083	0.076	0.060	0.043
0.3, 0.3, 0.3, 0.3	0.2, 0.2, 0.2, 0.2	0.064	0.062	0.055	0.048	0.064	0.063	0.053	0.049
0.3, 0.4, 0.5, 0.6	0.2, 0.2, 0.2, 0.2	0.060	0.057	0.051	0.045	0.056	0.059	0.054	0.048
0.3, 0.5, 0.6, 0.7	0.2, 0.2, 0.2, 0.2	0.056	0.054	0.048	0.046	0.054	0.052	0.053	0.044
0.5, 0.4, 0.3, 0.2	0.2, 0.2, 0.2, 0.2	0.068	0.069	0.062	0.053	0.068	0.069	0.056	0.048
0.3, 0.3, 0.3, 0.3	0.3, 0.3, 0.3, 0.3	0.058	0.056	0.053	0.049	0.057	0.059	0.052	0.054
0.3, 0.4, 0.5, 0.6	0.3, 0.3, 0.3, 0.3	0.051	0.054	0.044	0.044	0.048	0.050	0.049	0.047
0.3, 0.5, 0.6, 0.7	0.3, 0.3, 0.3, 0.3	0.048	0.050	0.047	0.041	0.042	0.046	0.044	0.040
0.5, 0.4, 0.3, 0.2	0.3, 0.3, 0.3, 0.3	0.060	0.066	0.059	0.053	0.066	0.065	0.056	0.048
0.3, 0.3, 0.3, 0.3	0.05, 0.1, 0.15, 0.2	0.738	0.786	0.698	0.665	0.742	0.791	0.702	0.674
0.3, 0.4, 0.5, 0.6	0.05, 0.1, 0.15, 0.2	0.697	0.688	0.659	0.632	0.691	0.696	0.668	0.644
0.3, 0.5, 0.6, 0.7	0.05, 0.1, 0.15, 0.2	0.651	0.662	0.639	0.613	0.669	0.659	0.641	0.614
0.6, 0.5, 0.4, 0.3	0.05, 0.1, 0.15, 0.2	0.692	0.695	0.635	0.586	0.708	0.710	0.653	0.597
0.3, 0.3, 0.3, 0.3	0.1, 0.25, 0.3, 0.3	0.878	0.869	0.866	0.845	0.864	0.875	0.860	0.847
0.3, 0.4, 0.5, 0.6	0.1, 0.25, 0.3, 0.3	0.841	0.844	0.835	0.826	0.837	0.839	0.828	0.821
0.3, 0.5, 0.6, 0.7	0.1, 0.25, 0.3, 0.3	0.826	0.827	0.819	0.808	0.816	0.825	0.808	0.808
0.6, 0.5, 0.4, 0.3	0.1, 0.25, 0.3, 0.3	0.828	0.826	0.797	0.793	0.825	0.828	0.803	0.777
0.3, 0.3, 0.3, 0.3	0.2, 0.2, 0.25, 0.4	0.590	0.599	0.577	0.556	0.627	0.624	0.598	0.592
0.3, 0.4, 0.5, 0.6	0.2, 0.2, 0.25, 0.4	0.513	0.520	0.509	0.501	0.547	0.561	0.533	0.533
0.3, 0.5, 0.6, 0.7	0.2, 0.2, 0.25, 0.4	0.482	0.497	0.478	0.462	0.510	0.527	0.510	0.488
0.6, 0.5, 0.4, 0.3	0.2, 0.2, 0.25, 0.4	0.553	0.562	0.551	0.514	0.594	0.597	0.567	0.552
0.3, 0.3, 0.3, 0.3	0.1, 0.2, 0.3, 0.4	0.959	0.958	0.954	0.953	0.965	0.964	0.963	0.958
0.3, 0.4, 0.5, 0.6	0.1, 0.2, 0.3, 0.4	0.939	0.940	0.937	0.936	0.945	0.945	0.943	0.932
0.3, 0.5, 0.6, 0.7	0.1, 0.2, 0.3, 0.4	0.927	0.934	0.923	0.920	0.922	0.935	0.921	0.923
0.6, 0.5, 0.4, 0.3	0.1, 0.2, 0.3, 0.4	0.942	0.940	0.928	0.922	0.951	0.954	0.937	0.930

Table 8: Simulated power of lower 95% confidence intervals of the Williams contrast and the Willimas contrast extended by a single linear contrast within settings with equal and increasing tumor rates and different mortality patterns

The results for the power simulation in Table 8 show,that in most cases the usage of the extended Williams contrast results in a little gain in power compared to the regular Williams contrast. In case of plateau shape in the true tumor incidence rates $(0.1, 0.25, 0.3, 0.3)$, the power for intervals based on the regular Williams contrast is in most cases higher. As already shown in Table 7, under the null hypothesis the Add-1 interval has coverage probablity nearest to the nominal level of 0.05.

References

[1] Armitage, P. (1955). Tests for linear trends in proportions and frequencies. *Biometrics* 11:375-386.

[2] Agresti, A., Caffo, B. (2000). Simple and effective confidence intervals for proportions and differences of proportions result from adding two successes and two failures. *American Statistician* 54:280-288.

[3] Bailer, J.A., Portier, C.J. (1988). Effects of treatment-induced mortality and tumor-induced mortality on tests for carcinogenicity in small samples. *Biometrics* 44:417-431.

[4] Bailer, J.A., Portier, C.J. (1989). Testing for increased carcinogenicity using a survival-adjusted quantal response test. *Fundamental and Applied Toxicology* 12:731-737.

[5] Bieler, G.S., and Williams, R.L. (1993). Ratio estimates, the Delta Method, and quantal response tests for increased carcinogenicity. *Biometrics* 49:793-801.

[6] Bretz, F. (2006). An extension of the Williams trend test to general unbalanced linear models. *Computational Statistics & Data Analysis* 50:1735-1748.

[7] Bretz, F., Hothorn, L.A. (2003). Statistical analysis of monotone or non-monotone dose response data from in vitro toxicological assays. *ATLA* 31:81-96.

[8] Bretz, F., Hothorn, L.A. (2002). Detecting dose-response using contrasts: asymptotic power and sample size determination for binomial data. *Statistics in Medicine* 21:3325-3335.

[9] Cochran W. (1954). Some methods for strengthening the common χ^2 tests. *Biometrics* 10:417-451.

[10] Dunnett, C.W. (1955). A multiple comparison procedure for comparing several treatments with a control. *Journal of the American Statistical Association* 50:1096-1121.

[11] Genz, A., Bretz, F., Hothorn, T. (2007). Mvtnorm: multivariate normal and t distribution. R package version 0.8-1.

[12] Hirotsu, C., Marumo, K. (2002). Changepoint Analysis as a Method for Isotonic Inference. *Scandinavian Journal of Statistics* 9:125-138.

[13] Hothorn, L.A., Bretz, F. (2000). Evaluation of animal carcinogenicity studies: Cochran armitage trend test vs. maximum tests. *Biometrical Journal* 42:553-567.

[14] Johnson, N.L., Kotz, S., Balakrishnan, N. (1994) *Wiley Series in Probability and Mathematical Statistics: Continuous univariate distributions.* New York: Wiley.

[15] Lin, K.K. (2000) Progress report on the guidance for industry for statistical aspects of the design, analysis and interpretation of chronic rodent carcinogenicity studies of pharmaceuticals.

[16] Moon, H., Hongshik, A., Kodell, R.L. (2006). A computational tool for testing dose-related trend using an age-adjusted bootstrap-based poly-k test. *Journal of Statistical Software* 16:Issue7.

[17] Sontag, J.M., Page, N.P., Saffiotti, U. (1976). Guidelines for carcinogen bioassay in small rodents. Technical Report 01; US Department of Health and Human Services: National Toxicology Program Research Triangle Park, NC.

[18] NTP) National Toxicology Program. (2000). Toxicology and carcinogenesis studies of methyleugenol in F344/N rats and B6C3F$_1$ mice.Technical Report 491. US Department of Health and Human Services: National Institutes of Health, Washington DC.

[19] Peddada, S.D., Dinse, G.E., Haseman, J.K. (2005). A survival-adjusted quantal response test for comparing tumor incidence rates. *Applied Statistics* 54:51-61.

[20] Peddada, S.D., Kissling, G.E. (2006). A survival-adjusted quantal response test for tumor incidence rates in animal carcinogenicity studies. *Environmental Health Perspectives* 114:537-541.

[21] Portier, C., Hedges, J., Hoel, D. (1986). Age-specific models of mortality and tumor onset for historical control animals in the National Toxicology Program's carcinogenicity experiments. *Cancer Research* 46:4372-4378.

[22] Price, R.M., Bonett, D.G. (2004). An improved confidence interval for a linear function of binomial proportions. *Computational Statistics and Data Analysis* 45:449-456.

[23] R Development Core Team (2007). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

[24] Ryan, T.P., Woodall W.H. (2005). The most-cited statistical papers. *Journal of Applied Statistics* 32:461-474.

[25] Williams, D.A. (1971). Test for differences between treatment means when several dose levels are compared with a zero control. *Biometrics* 27:103-117.