Leibniz
Universität Hannover

Reports of the Institute of Biostatistics

No 07 / 2007

Leibniz University of Hannover

Natural Sciences Faculty

Approximate simultaneous confidence intervals and tests for proportions using multiple Williams contrasts

Sill, M. Schaarschmidt, F.

1 Simultaneous confidence intervals for multiple contrasts

1.1 Approximate confidence intervals for a single linear combination of proportions

Let Y_i be independent binomial random variables $Y_i \sim Bin\left(n_i, \pi_i\right)$, $i = 1, ..., I$, with point estimators $p_i = Y_i/n_i$. Let $C = (c_1, ..., c_I)$ be a vector of contrast coefficients fulfilling the constraint $\sum_{i=1}^{I} c_i = 0$. Then the contrast is a linear combination $L = \sum_{i=1}^{I} c_i \pi_i$ which can be understood as difference of weighted averages of proportions, with the simple difference of two proportions as a special case. The point estimator for L is $\hat{L} = \sum_{i=1}^{I} c_i p_i$ and a Wald interval for L is:

$$\left[\sum_{i=1}^{I} c_i p_i \pm z_{1-\alpha/2} \sqrt{\sum_{i=1}^{I} c_i^2 \hat{V}\left(p_i\right)}\right] \tag{1}$$

with $\hat{V}\left(p_i\right) = p_i\left(1 - p_i\right)/n_i$. Wald intervals for binomial proportions are known to keep the $(1 - \alpha)$ coverage only for asymptotically large sample sizes. Even for single contrasts no exact confidence intervals are available. Improved Wald interval for the difference of two proportions by adding two successes and two failures has been proposed [14]. Bonett and Price [11], extended this interval to linear combinations of I proportions. In this approach, p_i in Equation (1) is replaced by $\tilde{p}_i = \frac{Y_i + 2/g}{n_i + 4/g}$, and $\hat{V}\left(p_i\right)$ by $\tilde{p}_i\left(1 - \tilde{p}_i\right)/\left(n_i + 4/g\right)$, with g the number of non-zero contrast coefficients. In another format of this method, p_i in Equation (1) is replaced by $\tilde{p}_i = \frac{Y_i + 1}{n_i + 2}$, and $\hat{V}\left(p_i\right)$ by $\tilde{p}_i\left(1 - \tilde{p}_i\right)/\left(n_i + 2\right)$. Both intervals are the Agresti and Caffo interval for the difference of two binomials if the contrast has only two non-zero coefficients. Price and Bonett [11], investigated the performance of their methods in a simulation study for different types of single contrasts. A better coverage for the improved intervals compared to the Wald interval has been revealed.

1.2 Approximate simultaneous confidence intervals for multiple contrasts

Interest lies in simultaneous estimation of confidence intervals for several, possibly correlated linear combinations of total order restricted proportions π_i. Then, we define M contrasts $C_m = (c_{m1}, \cdots, c_{mI})$, $m = 1, \cdots, M$ resulting in M linear combinations $L = (L_1, \cdots, L_M)$. Approximate simultaneous confidence intervals for the elements of L can be constructed.

$$\left[\sum_{i=1}^{I} c_{im} \tilde{p}_i \pm q_{M,R,1-\alpha} \sqrt{\sum_{i=1}^{I} c_{im}^2 \tilde{V}(p_i)}\right] \tag{2}$$

With respect to the approaches of Agresti and Caffo [14] and Price and Bonett [11], based on adding different numbers of successes and failures, several types of improved simultaneous intervals have been investigated. Table 1 summarizes parameters for $\tilde{p}_i$ and $\tilde{V}(p_i)$, where g is the number of non zero contrast coefficients in a single contrast. These parameters can be inserted in Equation 2, leading to approximate intervals. In the following these intervals are named corresponding to the notations in Table 1.

Notation	$\tilde{p}_i$	$\tilde{V}(p_i)$

Wald	Y_i/n_i	$p_i\left(1-p_i\right)/n_i$
add-1	$Y_i+0.5/n_i+1$	$\tilde{p}_i\left(1-\tilde{p}_i\right)/\left(n_i+1\right)$
add-2	Y_i+1/n_i+2	$\tilde{p}_i\left(1-\tilde{p}_i\right)/\left(n_i+2\right)$
add-2/g	$Y_i+\frac{1}{g}/n_i+\frac{2}{g}$	$\tilde{p}_i\left(1-\tilde{p}_i\right)/\left(n_i+\frac{2}{g}\right)$
add-4/g	$Y_i+\frac{2}{g}/n_i+\frac{4}{g}$	$\tilde{p}_i\left(1-\tilde{p}_i\right)/\left(n_i+\frac{4}{g}\right)$

Table 1: Parameters for $\tilde{p}_i$ and $\tilde{V}\left(p_i\right)$ in Equation 2

In Equation 2, $q_{M,R,1-\alpha}$ is the equicoordinate $(1-\alpha)$ quantile of a M-variate normal distribution with correlation matrix R with CDF $\Phi_M\left(\boldsymbol{q};\boldsymbol{0},\boldsymbol{R}\right)=P\left(|\boldsymbol{Z}|\leq q\right)$ for all elements of the M-variate normal random vector $\boldsymbol{Z}$. Most care implies total order restricted one-sided intervals, and therefore $\Phi_M\left(\boldsymbol{q};\boldsymbol{0},\boldsymbol{R}\right)=P\left(\boldsymbol{Z}\leq q\right)$ is used instead. Since the quantiles are chosen such that $P\left(|\boldsymbol{Z}|\leq q\right)=1-\alpha$ for all elements of $\boldsymbol{Z}$, the probability that at least one of the M values of $\boldsymbol{L}$ is excluded by the confidence intervals is α if $n\rightarrow\infty$.
Following Bretz and Hothorn [6] the correlation matrix R is

$$R=\begin{pmatrix}1&\rho_{12}&.&\rho_{1M}\\\rho_{21}&1&.&.\\.&.&1&.\\\rho_{M1}&.&.&1\end{pmatrix} \tag{3}$$

where

$$\rho_{mm`}=\frac{\sum_{i=1}^{I}c_{mi}c_{m`i}V(p_i)}{\sqrt{(\sum_{i=1}^{I}c_{mi}^2V(\,p_i))(\sum_{i=1}^{I}c_{m`i}^2V(p_i))}} \tag{4}$$

Further, following Piegorsch [17] we replace $V\left(p_i\right)$ by the parameter $\tilde{V}\left(p_i\right)$.
Notice, in the normal mode and under the assumption of equal variances the $\rho_{mm`}$s depend only on sample sizes n_i and the contrast coefficients c_{mi}.

1.3 Williams-type contrasts

In the examples presented, interest might be in a test of the null hypothesis of equal proportions $H_0:\pi_0=\pi_1=\cdots=\pi_I$ against an ordered alternative $H_1:\pi_0\leq\pi_1\leq\cdots\leq\pi_I$, with at least one strict inequality $\pi_1<\pi_i, i=2,\cdots,I$. For this purpose, one-sided multiple contrast tests ([9], [6]) or lower simultaneous confidence limits can be used. The contrasts are chosen such that the global null hypothesis of equality of all proportions is represented by $H_0:\bigcap_{m=1}^{M}L_m\leq 0$, while the alternative is $H_1:\bigcup_{m=1}^{M}L_m>0$, i.e. a Union-Intersection Test is performed for the M linear combinations. Bretz [9], showed that the Williams trend test can be expressed as multiple contrast test. Then, the $M=I-1$ contrasts are defined in the following matrix:

$$C_{M\times I}=\begin{pmatrix}-1&0&\cdots&0&0&1\\-1&0&\cdots&0&\frac{n_{I-1}}{n_{I-1}+n_I}&\frac{n_I}{n_{I-1}+n_I}\\\vdots&\vdots&\cdots&\vdots&\vdots&\vdots\\-1&\frac{n_2}{n_2+\cdots+n_I}&\cdots&\frac{n_{I-2}}{n_2+\cdots+n_I}&\frac{n_{I-1}}{n_2+\cdots+n_I}&\frac{n_I}{n_2+\cdots+n_I}\end{pmatrix} \tag{5}$$

In this multiple contrast the highest dose group is compared with the control, in the next single contrast the highest dose group and the next lower dose group are pooled, until all treatment groups are pooled and compared with the control. One can conclude for a trend in the proportions $\pi_1, \cdots, \pi_I$ if at least one of the M lower confidence limits excludes the value 0. Additionally to the decision of the global hypothesis, the confidence intervals display the difference between the control group $i = 1$ and the weighted average of proportions of higher dose groups.

2 Simulation study

We investigated the coverage probability of the simultaneous confidence intervals by a simulation study with focus on small to moderate sample sizes and $n_i = 100$ for nearly asymptotic behavior. The number of groups has been varied by $I = 3, 4, 6, 10$, with balanced sample sizes $n_i = 10, 20, 40, 60, 100$. To assess the methods performance for the whole parameter space, 10.000 combinations $\{\pi_1, \cdots, \pi_I\}$ were sampled from independent uniform distributions $[0, 1]$. For each of this combinations and settings, 10.000 random samples $\{y_1, \cdots, y_I\}$ were drawn from binomial distributions $Bin(n_i, \pi_i)$. The intervals were considered to cover the true value, if all estimated confidence intervals included their corresponding true linear combination L. For this main part of the simulation study the known values of π_i were used to calculate the correlation matrix, instead of using $\tilde{p}_i$. Further the coverage probability of the lower confidence bounds for some unbalanced situations has been investigated.
Additionally to the simulation study shown above, we simulated the coverage probabilities of intervals with the correlation matrix estimated from the samples. This was done for a small subset of situations to show whether the above simulations appropriately characterize the proposed methods. Like in the main study, for a balanced sample size of 40 and 4 groups, combinations $\{\pi_1, \cdots, \pi_I\}$ where drawn from uniform distribution and for each combination random samples have been generated. The coverage probability of each interval and method of computing the correlation was calculated, resulting in a negligible difference in coverage probability in the second position after the decimal point. To illustrate the oscillating behavior of the coverage probability for discrete intervals a third simulation study was performed. Here we investigated the coverage probability for 4 groups, one control group and three treatment groups. The π_is for the treatment groups were set to fixed value of 0.5 and 500 equally spaced values between 0 and 0.5 were chosen for π_0. A balanced sample size 50 was chosen. For each setting of π_0 again 10.000 simulation steps had been performed. We calculated the coverage probability of the lower confidence bounds for all investigated interval types. The results of this study are shown in Figure 1.

Taking the ideas of Agresti and Coull [15] and Brown and Li [16] into account, the coverage probability should be close to the nominal level. Therefore a proportion of coverage probability in a close range of let us say 94% to 96% around the nominal level is considered as main criterion for recommendation. This criterion is also sensible due to the oscillating behavior of the discrete confidence intervals as shown in Figure 1. Additionally, the mean coverage probabilities are given in brackets. The mean coverage probabilities give information wether the coverage probabilities lie mainly beyond or beneath the nominal level. The results for nominal 95%-confidence lower confidence limits are summarized in Table 2, while Table 3 displays the results for unbalanced situations.

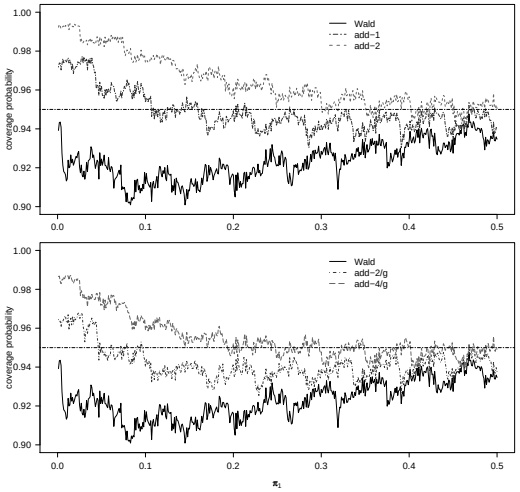

Figure 1: Coverage probability of the five investigated intervals for varying values of π_0 and fixed π_is of 0.5 for the treatment groups in a balanced four group design

I	n	Wald	add-1	add-2	add-2/g	add-4/g
3	10	0.208 *(0.913)*	0.506 *(0.951)*	0.234 *(0.959)*	0.481 *(0.948)*	0.315 *(0.957)*
3	20	0.359 *(0.933)*	0.700 *(0.950)*	0.359 *(0.955)*	0.676 *(0.949)*	0.463 *(0.954)*
3	40	0.502 *(0.942)*	0.832 *(0.950)*	0.492 *(0.953)*	0.827 *(0.949)*	0.617 *(0.952)*
3	60	0.600 *(0.945)*	0.884 *(0.950)*	0.576 *(0.952)*	0.880 *(0.949)*	0.702 *(0.952)*
3	100	0.722 *(0.947)*	0.931 *(0.950)*	0.688 *(0.951)*	0.933 *(0.950)*	0.797 *(0.951)*
4	10	0.202 *(0.908)*	0.487 *(0.951)*	0.269 *(0.959)*	0.385 *(0.944)*	0.433 *(0.954)*
4	20	0.320 *(0.930)*	0.674 *(0.950)*	0.394 *(0.956)*	0.540 *(0.946)*	0.598 *(0.952)*
4	40	0.436 *(0.941)*	0.814 *(0.950)*	0.541 *(0.953)*	0.717 *(0.948)*	0.748 *(0.951)*
4	60	0.519 *(0.944)*	0.875 *(0.950)*	0.632 *(0.952)*	0.813 *(0.948)*	0.813 *(0.951)*
4	100	0.641 *(0.946)*	0.925 *(0.950)*	0.740 *(0.951)*	0.900 *(0.949)*	0.882*(0.950)*
6	10	0.198 *(0.902)*	0.440 *(0.951)*	0.324 *(0.960)*	0.317 *(0.938)*	0.418 *(0.950)*
6	20	0.295 *(0.927)*	0.584 *(0.950)*	0.461 *(0.956)*	0.422 *(0.943)*	0.594 *(0.950)*
6	40	0.394 *(0.939)*	0.744 *(0.950)*	0.620 *(0.954)*	0.547 *(0.946)*	0.772 *(0.950)*
6	60	0.458 *(0.943)*	0.827 *(0.950)*	0.705 *(0.953)*	0.641 *(0.950)*	0.841 *(0.947)*
6	100	0.559 *(0.946)*	0.903 *(0.950)*	0.797 *(0.952)*	0.772 *(0.948)*	0.908 *(0.950)*
10	10	0.180 *(0.894)*	0.364 *(0.950)*	0.379 *(0.961)*	0.254 *(0.928)*	0.316 *(0.941)*
10	20	0.263 *(0.923)*	0.467 *(0.949)*	0.535 *(0.957)*	0.326 *(0.938)*	0.432 *(0.944)*
10	40	0.339 *(0.936)*	0.622 *(0.949)*	0.693 *(0.954)*	0.435 *(0.943)*	0.578 *(0.947)*
10	60	0.397 *(0.941)*	0.681 *(0.949)*	0.769 *(0.953)*	0.512 *(0.945)*	0.681 *(0.948)*
10	100	0.492 *(0.945)*	0.836 *(0.949)*	0.848 *(0.952)*	0.617 *(0.947)*	0.802 *(0.949)*

I	n	Wald	add-1	add-2	add-2/g	add-4/g

Table 2: Proportion of situations with coverage probability between 94% to 96% and mean coverage probability for one-sided 95% confidence intervals

In case of $3, 4$, or 6 groups, the add-1 and add-2/g interval achieve the highest proportions of coverage probabilities between 94% and 96%. For small sample sizes and large number of groups, these methods tend to be liberal. In these situations, the add-2 or add-4/g method are the better choice, if conservative performance is acceptable. As known, the Wald interval is more liberal than all other methods in all simulated situations. For the unbalanced four group designs, the add-2 and add-4/g

$n_1, \cdots, n_I$	Wald	add-1	add-2	add-2/g	add-4/g
64,32,32,32	0.513 *(0.941)*	0.864 *(0.950)*	0.509 *(0.953)*	0.862 *(0.948)*	0.723 *(0.952)*
80,40,30,10	0.280 *(0.910)*	0.503 *(0.947)*	0.467 *(0.955)*	0.474 *(0.945)*	0.664 *(0.954)*
10,30,40,80	0.147 *(0.903)*	0.312 *(0.951)*	0.608 *(0.961)*	0.241 *(0.941)*	0.497 *(0.952)*
20,30,50,60	0.228 *(0.930)*	0.494 *(0.950)*	0.645 *(0.956)*	0.367 *(0.945)*	0.736 *(0.951)*
60,50,30,20	0.382 *(0.934)*	0.800 *(0.950)*	0.495 *(0.953)*	0.765 *(0.947)*	0.717 *(0.952)*

Table 3: Proportion of situations with coverage probability between 94% to 96% and mean coverage probability for one-sided 95% confidence intervals in unbalanced situations

intervals approach the highest proportions of coverage probability between 94% and 96%. The add-1 intervals tends to be closest to the nominal level. Especially if the control group sample size is low, the perfomance of all intervals becomes weak.
If interest lies in two-sided confidence intervals, results of a simulation study for nominal 95% two-sided confidence intervals are summarized in Table 4.

I	n	Wald	add-1	add-2	add-2/g	add-4/g
3	10	0.001 *(0.878)*	0.410 *(0.947)*	0.573 *(0.960)*	0.263 *(0.941)*	0.595 *(0.957)*
3	20	0.007 *(0.917)*	0.738 *(0.947)*	0.746 *(0.956)*	0.628 *(0.944)*	0.794 *(0.953)*
3	40	0.342 *(0.935)*	0.937 *(0.948)*	0.883 *(0.953)*	0.926 *(0.947)*	0.910 *(0.952)*
3	60	0.692 *(0.940)*	0.973 *(0.949)*	0.938 *(0.952)*	0.973 *(0.948)*	0.954 *(0.951)*
3	100	0.902 *(0.944)*	0.990 *(0.949)*	0.971 *(0.951)*	0.992 *(0.949)*	0.976 *(0.951)*
4	10	0.000 *(0.869)*	0.287 *(0.946)*	0.532 *(0.961)*	0.129 *(0.934)*	0.432 *(0.952)*
4	20	0.001 *(0.913)*	0.637 *(0.946)*	0.702 *(0.956)*	0.338 *(0.940)*	0.764 *(0.951)*
4	40	0.196 *(0.932)*	0.929 *(0.948)*	0.859 *(0.953)*	0.860 *(0.944)*	0.921 *(0.950)*
4	60	0.567 *(0.944)*	0.973 *(0.950)*	0.919 *(0.952)*	0.958 *(0.948)*	0.962 *(0.951)*
4	100	0.866 *(0.943)*	0.993 *(0.949)*	0.969 *(0.951)*	0.992 *(0.948)*	0.983 *(0.950)*
6	10	0.000 *(0.857)*	0.166 *(0.944)*	0.505 *(0.961)*	0.075 *(0.922)*	0.090 *(0.943)*
6	20	0.000 *(0.906)*	0.463 *(0.945)*	0.671 *(0.957)*	0.107 *(0.934)*	0.479 *(0.945)*
6	40	0.117 *(0.929)*	0.888 *(0.947)*	0.810 *(0.954)*	0.584 *(0.941)*	0.901 *(0.947)*
6	60	0.425 *(0.936)*	0.963 *(0.948)*	0.883 *(0.953)*	0.865 *(0.944)*	0.961 *(0.948)*

I	n	Wald	add-1	add-2	add-2/g	add-4/g
6	100	0.762 *(0.942)*	0.989 *(0.948)*	0.951 *(0.952)*	0.980 *(0.946)*	0.987 *(0.949)*
10	10	0.000 *(0.842)*	0.099 *(0.942)*	0.474 *(0.961)*	0.049 *(0.904)*	0.040 *(0.927)*
10	20	0.000 *(0.898)*	0.259 *(0.943)*	0.667 *(0.957)*	0.047 *(0.924)*	0.094 *(0.936)*
10	40	0.075 *(0.925)*	0.796 *(0.946)*	0.797 *(0.954)*	0.341 *(0.936)*	0.617 *(0.942)*
10	60	0.347 *(0.933)*	0.927 *(0.947)*	0.856 *(0.953)*	0.619 *(0.941)*	0.844 *(0.945)*
10	100	0.656 *(0.940)*	0.980 *(0.948)*	0.918 *(0.952)*	0.864 *(0.944)*	0.972 *(0.947)*

Table 4: Proportion of situations with coverage probability between 94% to 96% and mean coverage probability for two-sided 95% confidence intervals

Here the highest proportion of coverage probability between 94% and 96% is achieved by the two-sided add-2 interval. This interval is more conservative, but for small sample sizes its coverage probability is closest to the nominal level. The most liberal interval is again the Wald, which has in all simulated settings lower coverage probability compared to the other methods. The add-2/g and the add-4/g method perform comparable to the add-1 and add-2 for situations with 3 and 4 groups. With an increasing number of groups both methods get more liberal.

3 Approximative power calculation

Unless our main focus is confidence interval estimation, users might be interested in power calculation for the proposed global test. Bretz and Hothorn [6] derive an approximative calculation for the any-pairs power of contrast tests for binary data. Their approach differs from the above version by using the maximum likelihood estimators for variance estimation instead of the adjustments proposed in this paper, and by using a pooled variance estimator under the null hypothesis. Under the alternative, we assume true proportions π_i and sample sizes n_i. Then, for large n_i, a single test statistic follows a normal distribution with expectation

$$E\left(T_m\right) = \frac{\sum_{i=1}^{I} c_{im}\tilde{\pi}_i^*}{\sqrt{\sum_{i=1}^{I} c_{im}^2 \tilde{\pi}_i^* \left(1-\tilde{\pi}_i^*\right)/\tilde{n}_i^*}} \tag{6}$$

and variance $V\left(T_m\right) = 1$, where $\tilde{\pi}_i^* = \left(n_i\pi_i + 0.5\right)/\left(n_i+1\right)$ and $n_i^* = n_i + 1$, e.g. for the add-1 adjustment. The M test statistics jointly follow an M-variate normal distribution with $\mathbf{e}$ the vector of expectations with elements $E\left(T_m\right)$, and correlation matrix $\mathbf{R}$ as defined in Equations (3), and (4). The power to reject the global null hypothesis is the probability that at least one T_m exceeds the equicoordinate critical value $q_{M,R,1-\alpha}$. Therefore, it can be calculated using: $1 - \Phi_M\left(\mathbf{q}_{\mathbf{M},\mathbf{R},\mathbf{1}-\alpha}; \mathbf{e}, \mathbf{R}\right)$ or equivalently by using a central multivariate normal distribution after subtracting the vector of expected values from the quantiles: $1-\Phi_M\left(\mathbf{q}_{\mathbf{M},\mathbf{R},\mathbf{1}-\alpha} - \mathbf{e}; \mathbf{0}, \mathbf{R}\right)$.Clearly, this power calculation gives misleading results in situations where the normal approximation fails, i.e. for small sample sizes and extreme small and large proportions. Note further, that due to the discreteness of the binomial distribution, the true power in dependence of n_i, π_i is a non-monotone function and thus necessarily deviates from any approximation by monotone functions. A simulation study was performed for 150 different settings of

increasing ($H_1 : \bigcup_{m=1}^{M} L_m > 0$) and decreasing ($H_2 : \bigcup_{m=1}^{M} L_m < 0$) trends with $0.05 \leq \pi_1 \leq 0.95$, balanced sample sizes and 10.000 simulation steps for each setting. To summarize the results, the absolute difference between calculated and simulated power is presented. In Table 5 the percentage of settings for which the absolute difference is smaller than 0.02 is shown, additionally the maximal absolute difference is given in italics. The maximal absolute differences usually occurred for $\pi_1 = 0.95$ with H_{11} and $\pi_1 = 0.05$ with H_{12}. Thus for sample sizes smaller than 40 and expected proportions close to 0 or 1, simulation of power is recommended instead of using the approximation above.

In Table 6, approximate power calculation is compared to simulated power (10.000 replications) for

ni	H_1, Wald	H_1, add-1	H_1, add-2	H_2, Wald	H_2, add-1	H_2, add-2
20	0.527 *(0.163)*	0.533 *(0.148)*	0.527 *(0.125)*	0.547 *(0.164)*	0.527 *(0.146)*	0.54 *(0.125)*
40	0.707 *(0.196)*	0.773 *(0.153)*	0.720 *(0.159)*	0.693 *(0.198)*	0.780 *(0.156)*	0.713 *(0.163)*
60	0.780 *(0.125)*	0.793 *(0.096)*	0.767 *(0.156)*	0.773 *(0.121)*	0.793 *(0.100)*	0.800 *(0.153)*
100	0.887 *(0.045)*	0.847 *(0.058)*	0.827 *(0.067)*	0.880 *(0.043)*	0.853 *(0.052)*	0.833 *(0.061)*
200	0.940 *(0.081)*	0.940 *(0.072)*	0.92 *(0.074)*	0.933 *(0.080)*	0.947 *(0.072)*	0.92 *(0.075)*

Table 5: Absolute difference between approximative and simulated power: percentage of 150 settings with absolute difference ≤ 0.02, and maximal absolute difference in italics

tests on increasing trend with nominal level $\alpha = 0.05$, using the add-1 adjustment. Three different dose response shapes are assumed, which could be underlying the data in Table **??**; power is calculated for balanced samples sizes $n_i = 30, 40, 50$.

π_1	π_2	π_3	π_4	n_i	Approximate power	Simulated power
0.15	0.15	0.15	0.4	30	0.6263	0.6436
0.15	0.15	0.15	0.4	40	0.7485	0.7586
0.15	0.15	0.15	0.4	50	0.8371	0.8452
0.1	0.1	0.15	0.4	30	0.8199	0.8229
0.1	0.1	0.15	0.4	40	0.9134	0.9130
0.1	0.1	0.15	0.4	50	0.9603	0.9576
0.1	0.25	0.25	0.25	30	0.5924	0.5951
0.1	0.25	0.25	0.25	40	0.7117	0.7036
0.1	0.25	0.25	0.25	50	0.8007	0.7911

Table 6: Approximate and simulated power of tests for increasing trend ($\alpha = 0.05$) using add-1 adjustment

Note, that for such small sample sizes the approximate power calculation might show larger deviations from true power than the examples in Table 6 reveal. However, taking the uncertainty in assuming $\pi_1, \cdots, \pi_I$ into account, it can be considered as a helpfull tool in experimental design for moderate sample sizes.

References

[1] Williams DA. Test for differences between treatment means when several dose levels are compared with a zero control. *Biometrics* 1971; 27: 103-117.

[2] Clayton NP, Yoshizawa K, Kissling GE. Immunohistochemical analysis of expressions of hepatic cytochrome P450 in F344 rats following oral treatment with kava extract. *Exp Toxicol Pathol.* 2007; 58: 223–236.

[3] Stewart WH, Ruberg SJ. Detecting dose response with contrasts. *Statistics in Medicine* 2000; 19: 913-921.

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

[5] Armitage P. Tests for linear trends in proportions and frequencies. *Biometrics* 1955; 11: 375-386.

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

[7] Tang ML, Ng HKT, Guo JH, Chan W, Chan BPS. Exact Cochran-Armitage trend tests: comparisons under different models. *Journal of Statistical Computation and Simulation* 2006; 76: 847-859.

[8] Leuraud K, Benichou J. A comparison of stratified and adjusted trend tests for binomial proportions. *Statistics in Medicine* 2006; 25: 529-535.

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

[10] Roehmel J. Problems with existing procedures to calculate exact unconditional p-values for non-inferiority/superiority and confidence intervals for two binomials and how to resolve them. *Biometrical Journal* 2004; 47: 37-47.

[11] Price RM, Bonett DG. An improved confidence interval for a linear function of binomial proportions. *Computational Statistics & Data Analysis* 2004; 45: 449-456.

[12] Hothorn L. Multiple comparisons in long-term toxicity studies. *Environmental Health Perspectives* 1994; 1: 33-38.

[13] Berth-Jones J, Todd G, Hutchinson PE, Thestrup-Pedersen K, Vanhoutte FP. Treatment of psoriasis with oral liarozole: a dose-ranging study. *British Journal of Dermatology* 2000; 143: 1170-1176.

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

[15] Agresti A, and Coull A. Approximate is better than "exact" for interval estimation of binomial proportions. *American Statistician* 1998; 52: 119-126.

[16] Brown L, Li X. Confidence intervals for two sample binomial distribution. *Journal of Statistical Planning and Inference* 2003; 130: 359-375.

[17] Piegorsch WW. Multiple comparisons for analyzing dichotomous response. *Biometrics* 1991; 47: 45-52.