Ulm, Dannegger, Becker:

Tests for Trends in Binary Response

Sonderforschungsbereich 386, Paper 115 (1998)

Online unter: http://epub.ub.uni-muenchen.de/

Tests for Trends in Binary Response

KURT ULM
FELIX DANNEGGER
URSULA BECKER
Institut für Medizinische Statistik und Epidemiologie
Technische Universität München
Ismaninger Straße 22, 81675 München, Germany
email: kurt.ulm@imse.med.tu-muenchen.de

Tests for Trends in Binary Response

SUMMARY

Tests for trend in binary response are especially important when analyzing animal experiments where the response in various dose–groups is of interest. Among the nonparametric tests the approach of Cochran and Armitage is the one which is most commonly used. This test (CA-test) is actually a test for a linear trend. The result of this test is highly dependent on the quantification of the dose. Varying score assignments can lead to totally different results. As an alternative isotonic regression is proposed. The result of this approach is independent of any monotonic transformation of the dose. The p–value related with the isotonic regression can be obtained either from considering all possible combinations of the total number of events in the dose–groups or by analyzing a random sample of all permutations. Both tests are compared within a simulation–study and on data from an experiment considering whether a certain type of fibre, para–aramid, is carcinogenic. The result of the commonly used CA–test is highly dependent on the event rate in the lowest and highest dose–group. Based on our analyses we recommend to use the isotonic regression instead of the test proposed by Cochran and Armitage.

1. INTRODUCTION

Many decisions whether a certain agent is carcinogenic or not are based on animal experiments. The list of carcinogenic compounds in the work area in Germany for example currently contains over 100 substances with only 21 having been classified on the basis of epidemiological data.[1] For all other decisions only data from animal experiments are available. One main criterion in order to establish causality is the proof of a dose–response relationship. In order to simplify the analysis only the situation of a binary outcome (e. g. tumor yes or no) is considered ignoring the time to the event. This is common practice in analysing animal experiments. Very recently a review of tests available for this situation was published by Chuang-Stein and Agresti.[2] A variety of tests are described, parametric as well as nonparametric ones. Among the nonparametric tests mentioned in this tutorial, the so called Cochran–Armitage test (CA–test) seems to be the most commonly used one.[3,4] This test is actually a test for a linear trend. The main problem associated with this test is the necessity of assigning scores to the various dose–levels. This potential disadvantage was also mentioned by Chuang-Stein and Agresti.[2] They recommend to use to actual dose level or the log dose. Another approach is to use the index i. They report that the choice has little impact on the results, except when the data are highly unbalanced. We want to present an actual example where the conclusion is highly dependent on the assigned scores.

The classification of man–made mineral fibres as carcinogenic (or not) is still controversial. Especially one type of fibre, para-aramid, is heavily under discussion.[5] The main source of data to classify this type of fibre is an animal experiment which has been evaluated many times. The latest update is given in table 1.

Table 1: Data from the para–aramid study.

dose $[*10^6 F/m^3]$	0	2.5	25	100	400
number of tumors*	1	1	1	4	3
number of animals	137	133	132	137	92

* Adenoma, bronchido-alveolar without keratinizing squamous-cell carcinoma

Applying the CA–test using various choices to assign scores to the dose levels leads to the following results (one–sided tests). The test–statistic (χ^2_{CA}) is assumed to be χ^2 distributed with one degree of freedom. The critical value for a significance level of 5% for the one sided test will be 2.71 ($\chi^2_{1,90\%}$). Two

of the three assignments lead to a statistically significant result, whereas the third one, using log dose, yields a non–significant result. The discussion about the classification of para–aramid fibres is still controversial and the differences in the test results are not helpful in that discussion.

The result of the CA–test at least in this example is highly dependent on the choice of the scores comparable to linear regression analysis where a transformation of the x–axis can yield different results. In contrast, a test based on the result of a monotonic analysis, the isotonic regression can be applied. The result is independent of the score assignment to the various dose–groups. The distribution of the test–statistics for the isotonic regression for large samples is well–known.[6] Within this paper the CA–test will be compared with the isotonic regression. First a simulation study is performed where several conditions are investigated. Afterwards the para–aramid data are considered in greater detail.

2. METHODS

2.1. Description of the data

The data of the type of experiments which are considered throughout the paper can be displayed as in table 2.

Table 2: Notation used for the various test–statistics.

dose–level	d_0	d_1	...	d_k	$\sum$
number of tumors	r_0	r_1	...	r_k	r
number of subjects	n_0	n_1	...	n_k	n
proportion	$p_0 = \frac{r_0}{n_0}$	$p_1 = \frac{r_1}{n_1}$	...	$p_k = \frac{r_k}{n_k}$	$\bar{p} = \frac{r}{n}$
score	x_0	x_1	...	x_k	

2.2. Cochran–Armitage test (CA–test)

If the proportion of events under dose d_i is denoted with p_i, the assumption

$$p_i = \alpha + \beta x_i \tag{1}$$

is hypothesized. In (1), β is the slope of the regression line, which we are interested in. β can be estimated by the usual least squares method,[7] with

$$\hat{\beta} = \frac{\sum n_i (p_i - \bar{p})(x_i - \bar{x})}{\sum n_i (x_i - \bar{x})^2}.$$

The test due to Cochran[3] and Armitage[4] is used to check whether the slope β is indeed different from 0. The usual chi-square test

$$\chi^2 = \sum_{i=0}^{k} \left\{ [r_i - \mathrm{E}\,(r_i)]^2 \left[\frac{1}{\mathrm{E}\,(r_i)} + \frac{1}{n_i - \mathrm{E}\,(r_i)} \right] \right\},$$

$$\text{with} \quad \mathrm{E}\,(r_i) = \frac{r}{n} n_i$$

checking the association between the dose and the rate of tumors can be decomposed into two parts

$$\begin{aligned} \chi^2 &= \chi^2_{linearity} + \chi^2_{slope} \\ \chi^2_{linearity} &= \frac{\sum n_i (p_i - \hat{p}_i)}{\bar{p}(1-\bar{p})} \end{aligned}$$

with $\hat{p_i}$ being the estimated proportion corresponding to the linear regression in (1). The statistic $\chi^2_{linearity}$ has $k-2$ degrees of freedom and the hypothesis of linearity would be rejected if $\chi^2_{linearity}$ were found to be too large. The statistic

$$\chi^2_{slope} = \hat{\beta}^2 \sum \frac{n_i (x_i - \bar{x})^2}{\bar{p}(1-\bar{p})}, \tag{2}$$

usually called the Cochran–Armitage test (CA–test) has 1 degree of freedom and may be used to test the significance of the slope ($H_0 : \beta = 0$).

The main problem is related to the characterization of the dose d_i, which is used as x_i in estimating β and calculating (2). A small example may demonstrate the effect. We will consider 4 dose groups with $d_0 = 0$, $d_1 = 2.5$, $d_2 = 25$ and $d_3 = 250$ mg with 50 animals in each group. The number of tumors in the four groups are assumed to be $(0, 4, 5, 6)$. If the dose groups are assigned with their actual dose–levels, the test–statistic leads to a value of $\chi^2_{slope} = 2.32$ ($\hat{\beta} = 0.0003$) which results in a p–value of $p = 0.10$. Assigning the index to the dose–groups ($x_i = 0, 1, 2, 3$), the associated test–statistic $\chi^2_{slope} = 5.20$ ($\hat{\beta} = 0.038$) indicates a significant slope ($p = 0.022$). Using $\log\,(dose + 0.01)$ as scores the corresponding test–statistic χ^2_{slope} is 5.88 ($\hat{\beta} = 0.028$) which is statistically significant ($p = 0.016$). The overall χ^2–test gives a value of 5.98. This indicates that $\chi^2_{linearity}$ is close to zero for the index and the log dose score assignment. However if the doses are used as scores the value for $\chi^2_{linearity}$ of 3.66 would be too low to reject the hypothesis of linearity.

This result is puzzling and an approach is needed which gives a result independent of the choice of dose–assignment.

2.3. Isotonic Regression

Isotonic regression theory provides a nonparametric solution to the problem of monotonicity.[6] The solution maximizes the likelihood function under the

constraints of monotonicity

$$p_0 \leq p_1 \leq \cdots \leq p_k.$$

If this relation in the observed proportions is not fulfilled for one neighboring pair $(i, i+1)$ both groups are pooled using

$$p_i^* = p_{i+1}^* = \frac{r_i + r_{i+1}}{n_i + n_{i+1}}$$

in order to give the same averaged response rate to both groups. When repeatedly used until all neighboring groups adhere to the monotonicity constraint this procedure is called the 'pool adjacent violators algorithm' by Robertson, Wright and Dykstra.[6] The test–statistic in order to check a dose–response relationship is based on the likelihood ratio statistic R comparing the likelihood function under H_0 (the response rates in all dose–groups are equal) and under the alternative H_1 (the result of the isotonic regression), where at least one inequality strictly holds. With the notation used in table 2 the likelihood functions are as follows

$$\begin{aligned} \ln \mathrm{L}\,(\mathrm{H}_0) = \ln \mathrm{L}_0 &= \sum_{i=0}^{k} [r_i \ln \bar{p} + (n_i - r_i) \ln (1 - \bar{p})] \\ &= r \ln \bar{p} + (n - r) \ln (1 - \bar{p}), \\ \text{with} \quad \bar{p} &= \frac{r}{n} \end{aligned}$$

being the overall event rate. Under the alternative H_1 the value of the likelihood function is

$$\ln \mathrm{L}\,(\mathrm{H}_1) = \ln \mathrm{L}_1 = \sum_{i=0}^{k} [r_i \ln p_i^* + (n_i - r_i) \ln (1 - p_i^*)]$$

with p_i^* being the maximum likelihood estimates under the constraints of monotonicity. The test–statistic in order to prove H_1 against H_0 is the likelihood ratio test

$$\mathrm{R} = 2 * \{\ln \mathrm{L}_1 - \ln \mathrm{L}_0\}.$$

The distribution of the test–statistic R under H_0 is

$$\Pr(R > c) = \sum_{j=2}^{k} \left[\Pr\left(\chi_{j-1}^2 > c\right) w(j,k)\right]$$

with $w(j,k)$ $(\sum_{j=1}^{k} w(j,k) = 1)$ denoting the probabilities that given k subgroups with equal number of animals under H_0 the isotonic regression will result in j different estimates of p_j.

The example considered in the previous chapter with 4 dose groups, 50 animals per group and an outcome of $(0, 4, 5, 6)$ leads to a value of $R = 9.48$ which is statistically significant $(p < 0.01)$. The behaviour of the test–statistics for the isotonic regression will be investigated in the next chapter.

3. RESULTS OF A SIMULATION STUDY

First of all the behaviour of the various tests is investigated in large and small samples. In analogy to the para–aramid data the situation with 5 dose groups and 50 animals within each dose–group is considered. Each simulation is analyzed with four different approaches, the CA–test assigning the dose, the index or $\log(dose + 0.01)$ to the dose–groups and isotonic regression. All results are based on 10000 replications. Simulations were performed with differing event rates and assuming equal risks in each group.

3.1. Event rate = 10%

The proportion of events is assumed to be 10%. On average 5 out of 50 animals in each group will develop the disease. The distribution of the CA–test assigning the different scores to the dose–groups is displayed in figure 1 together with the χ^2_1 distribution. There is nearly no difference between the four curves. From this simulation one can conclude at least two aspects. First, under H_0 all three assignments lead to the same conclusion. The distributions of all three strategies are nearly identical. Second, the empirical distributions approximately follow a χ^2 distribution with one degree of freedom.

Applying isotonic regression, the empirical distribution of the test–statistics is shown in figure 2 together with the theoretical one. Again there is good agreement between both curves.

3.2. Event rate = 4%

The same situation but with less events was considered. On average 2 out of the 50 animals within each of the five dose–groups will develop the disease. In figure 3 the results of the CA–test with the three different assignment of scores to the dose–groups are shown together with the χ^2 distribution. Even in this situation there is good agreement between all three curves and the χ^2 distribution with 1 degree of freedom.

If the number of events is even smaller ($p = 2\%$), the agreement between the empirical and the theoretical distribution remains fairly good. Applying isotonic regression a somewhat different result, depicted in figure 4 is obtained. The empirical distribution is different from the theoretical one. As an example, consider the critical value for testing H_0 at the 5% which has changed from 5.05 to 5.71.

The conclusion from this analysis is, if the number of events is small, the test based on the critical value e. g. from table 4 out of the monograph of

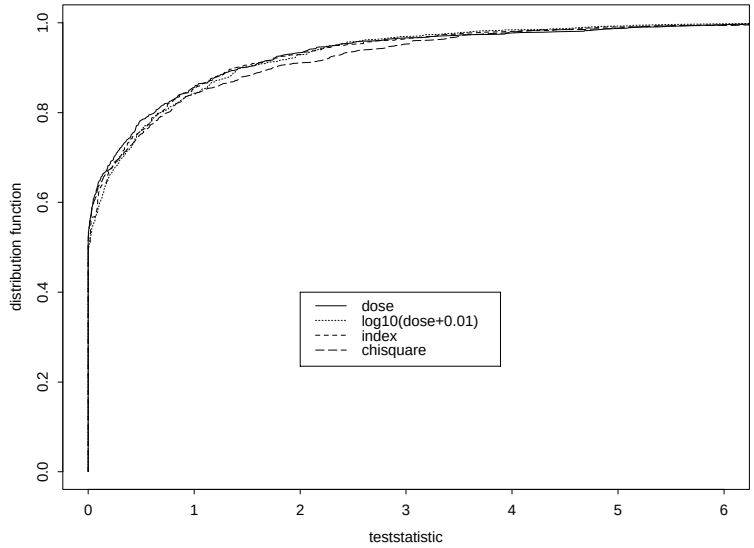

Figure 1: Distribution of the three CA–tests using different score assignments under H_0 with an overall event rate of $\bar{p} = 10\%$ considering 5 dose groups with 50 animals each. The results are based on 1000 simulations. The χ^2- distribution with 1 degree of freedom is also plotted as a guide.

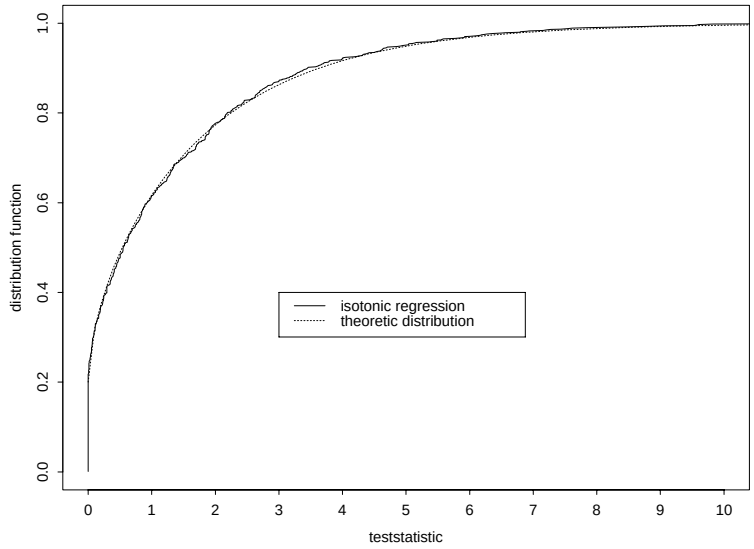

Figure 2: Distribution of the observed likelihood ratio statistics R for isotonic regression together with the distribution function obtained from the large sample approximation. The same situation as in figure 1 is considered.

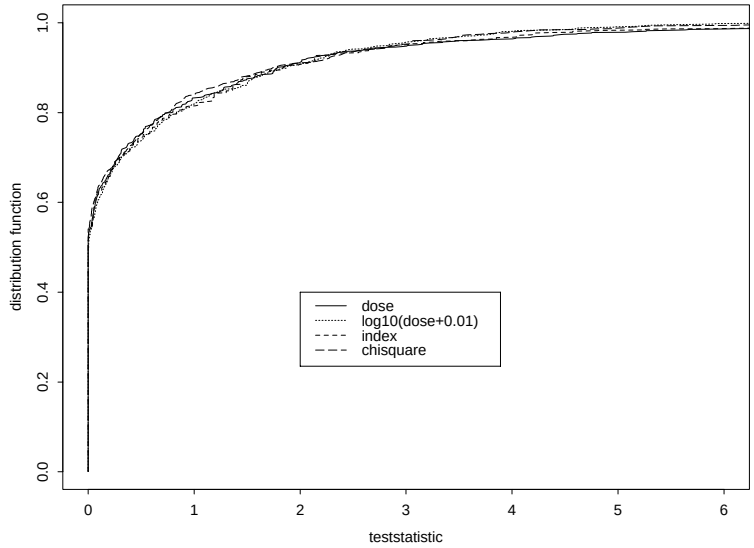

Figure 3: Distribution of the three CA–tests using different score assignments under H_0 with an overall event rate of $\bar{p} = 4\%$ considering 5 dose groups with 50 animals each. The results are based on 1000 simulations. The χ^2- distribution with 1 degree of freedom is also plotted as a guide.

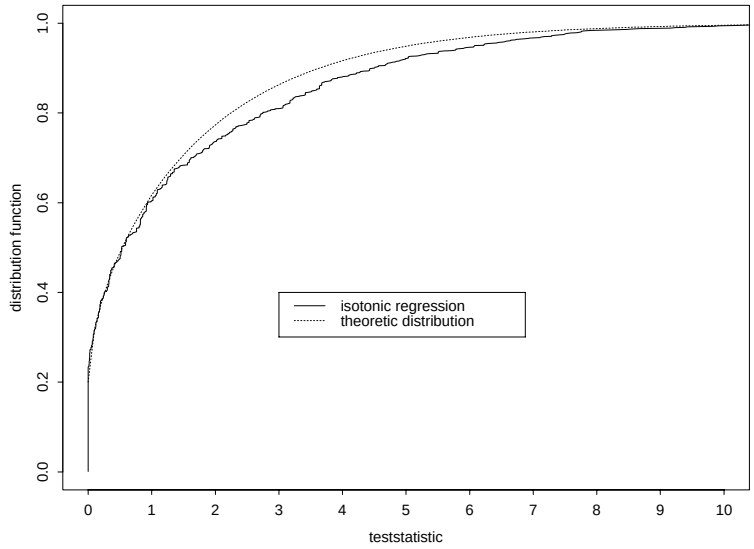

Figure 4: Distribution of the observed likelihood ratio statistics R for isotonic regression together with the distribution function obtained from the large sample approximation. The same situation as in figure 3 is considered.

Robertson et al.[6] is misleading. Therefore the decision to reject H_0 cannot be based on the values derived from the large sample approximation. From these simulations one can conclude that the critical value is higher than the tabulated one. The decision about rejecting or accepting H_0 has to be based on additional analyses.

In this situation performing a permutation test is recommended Chen, Kodell and Pearce.[8] This test is described in the next section. But before that, the power of the various tests is investigated. The dose–levels and the number of animals per dose–level ($n = 50$) are identical to the situation under H_0.

Several different situations were investigated: In the various situations one of the three score assignments leads to a perfectly linear relationship. In the CA–test the hypothesis we are interested in is $H_0 : \beta \leq 0$ versus $H_1 : \beta > 0$. The significance level used is $\alpha = 5\%$. In this one–sided test-situation H_0 is to be rejected if the test–statistics χ^2_{slope} exceeds 2.71 ($\chi^2_{1,90\%}$) and the estimate for β is positive. The probabilities of rejecting H_0 for the three score assignments are displayed in table 3.

The various situations were also analyzed using isotonic regression. If the overall event rate is 10% the hypothesis H_0 is rejected if the test–statistic R exceeds 5.05 ($\alpha = 0.05$). In situations where the event rate is 4%, the critical value obtained from the simulation study ($R = 5.71$, $\alpha = 0.05$) is taken. The results of the isotonic regression are also given in table 3.

In most situations there is fairly good agreement in the results for power. In some situations however the difference can be in the order of 20% or even more. Especially the use of the dose as score assignment can lead to poor power. In cases where the event rate in the higher dose–groups are nearly identical, the assumption of linearity with respect to dose is violated. The differences in the power in using either log dose or the index are much smaller, less than 7% in the situations we considered. On the other hand, isotonic regression performs much better. Overall isotonic regression gives the best values. The difference to the result of the CA–test based on the assignment with a perfect linear fit is fairly small. In the situations considered the differences are less than 2.5%. In order to make a more general recommendation, the mean over all nine situations can be considered. Overall power is 61.6% for isotonic regression, followed by 61.3% using the index as scores. The log dose assignment leads to average power 60.6%. The lowest power of 56.9% is observed using the dose as scores. From this analysis the use of the isotonic regression can be recommended.

Table 3: Probability of rejecting H_0 (power) for the various situations considered by the CA–test and isotonic regression. Five groups with 50 animals each with the proportion of events being $(p_1, \ldots, p_5)$ and the overall rate $\bar{p}$ are investigated. The critical values c assuming a significance level of $\alpha = 5\%$ are as follows: $c = 2.71$ for the CA–test, $c = 5.05$ using isotonic regression and assuming $\bar{p} = 10\%$ and $c = 5.71$ using isotonic regression and assuming $\bar{p} = 4\%$.

			Power %			
$(p_1, \ldots, p_5)$	$\bar{p}$	linear	CA–test with score			isotonic
(%)	(%)	relationship with	dose	index	log dose	regression
(2, 6, 10, 14, 18)	10	index	84.8	94.5	94.9	93.7
(6, 8, 10, 12, 14)	10	index	41.3	45.6	43.3	42.4
(2, 3, 4, 5, 6)	4	index	30.8	31.2	28.3	29.8
(4.8, 4.9, 5.8, 14.9, 19.8)	10	dose	94.0	92.4	86.7	90.4
(7.4, 7.4, 7.9, 12.4, 14.9)	10	dose	49.8	43.6	36.9	41.1
(0.4, 0.4, 1.1, 7.4, 10.8)	4	dose	97.4	96.6	95.0	95.6
(2.1, 8.1, 11.1, 14.1, 14.7)	10	log dose	60.3	78.9	84.8	83.9
(8.4, 9.6, 10.2, 10.8, 10.9)	10	log dose	10.8	11.3	10.6	11.2
(0.1, 3.1, 4.6, 6.1, 6.3)	4	log dose	42.7	57.3	65.1	66.4
		mean	56.9	61.3	60.6	61.6

4. PERMUTATION TESTS

Based on the results from the simulation studies it is advised to give the p–value not based on the theoretical distribution, especially if the event–rate is fairly small. In this situation an exact method is to be preferred. One way to estimate the correct p-value is to perform a permutation–test. Based on the observed margins (number of animals per dose–group and total number of events) a large number of permutations (e.g. $M = 10000$) are analyzed. Each animal is characterized by a pair $((d_i, \delta_i)$ with $i = 1, \ldots, n)$ with d_i denoting the dose–group and δ_i the status, taking $\delta_i = 1$ to indicate the occurence of an event and $\delta_i = 0$ otherwise.

For the permutation–test this pair is broken up and dose–level and status components are combined on random allocation. Within each permutation H_0 is considered (H_0: equal risk in all dose–groups). Each permutation is analyzed by the test proposed. If T_{obs} is the observed value of the test–statistic of the original data the estimated p–value is merely the estimate of the probability

that the result of a permutation is equal to T_{obs} or exceeds it. Formally

$$\hat{p} = \widehat{\Pr}(T_{perm} \geq T_{obs}) = \sum_{m=0}^{M} I_{\{T_{perm}^{(m)} \geq T_{obs}\}},$$

with $T_{perm}^{(m)}$ denoting the m-th observed permuted test–statistic. If $\hat{p}$ is less than the predefined significance-level, H_0 is to be rejected and a dose–response relationship can be assumed.

The alternative to the permutation test is to look at all possible combinations, applying the test proposed to all these combinations, calculating the probabilities for the combinations with a test–statistic equal or greater as the observed one and adding all these probabilities up. If the sum is less than the predefined significance level, H_0 is to be rejected. The probability of observing the combination $(r_1, r_2, r_3, r_4, r_5)$ with $\sum r_i = r$ is

$$p(r_1, \ldots, r_5) = \frac{\binom{n_1}{r_1}\binom{n_2}{r_2}\cdots\binom{n_5}{r_5}}{\binom{n}{r}}.$$

The number of possible combinations depends on the total number of different dose–groups. In the para–aramid example with $r = 10$ events and 5 dose–groups a total of 1001 combinations are possible. The exact p–value is the sum of all probabilities of those combinations with a test–statistics equal or greater than the observed one.

5. APPLICATION TO THE PARA–ARAMID DATA

The data presented in table 1 are analyzed with the various tests in greater detail. The results are summarized in table 4. For the CA–test the results of both test statistics χ^2_{slope} and $\chi^2_{linearity}$ as well as the estimates of the slope β are given.

The decision concerning the acceptance of H_0 depends on the test used as well as on the method the p–value is obtained. If the CA–test is applied H_0 will be rejected if the indices are used as scores. The test–statistic χ^2_{slope} of 3.819 is statistically significant in all cases based on the χ^2 distribution ($p = 0.026$), using permutations ($\hat{p} = 0.039$) and analyzing all 1001 combinations ($p = 0.033$). If the other two methods of assigning the scores are used, the p–values are sometimes below and sometimes above 0.05. The result of the CA–test is highly dependent on the way the test is performed. On the other hand the isotonic regression leads to a different conclusion. The hypothesis H_0 cannot be rejected. The p–value obtained from the large sample approximation is slightly above 0.05 ($p = 0.058$). The p–value based on a sample of 10000

Table 4: Results of the analyses of the para–aramid data from table 1.

	CA–test			
	score assignment			isotonic
test–statistics	dose	index	log (dose + .01)	regression
χ^2_{slope} or R	3.294	3.819	2.533	4.78
$\chi^2_{linearity}$	1.741	1.216	2.502	
slope β	$7 \cdot 10^{-5}$	0.008	0.005	
p-value of χ^2_{slope} or R				
tabulated	0.035	0.026	0.056	0.058
permutation test ($M = 10000$)	0.049	0.033	0.047	0.110
exact	0.049	0.033	0.048	0.111

randomly selected permutations and on all possible combinations (exact) are both above 0.05 ($p = 0.11$). This result is in line with the results from the simulation study. If the event rate is small, the p–values obtained from the large sample approximation are misleading. The exact p–value as well as the p–value from the permutation test show no significant dose–response relationship. In order to investigate these differences in more detail some results of the isotonic regression and the CA–test using the index as scores are considered. In figure 5 the results of both tests for all 1001 combinations are shown. In general there is good agreement between both tests.

However some of the combinations have totally different outcomes. For example the permutation $(0, 0, 6, 3, 1)$ leads to a nonsignificant value of $\chi^2_{slope} = 2,21$ ($\hat{\beta} = 0.006$) using the CA–test. The isotonic regression yields a value of $R = 11.29$ ($p < 0.01$). There is also a difference to the two other methods of assigning scores. If the dose is used, the slope switches signs and turns negative, whereas the log (dose)–method gives a test value of $\chi^2_{slope} = 3.48$. The other situations where both tests are different are similar. The proportion of events follows more or less an umbrella– or u–shape. The risk is high in the dose groups in the middle and low at both ends or vica versa. Isotonic regression amalgamates the highest or lowest dose group together with the dose– groups in the middle which leads to an increased risk in the higher dose groups. Regression analysis however leads to a more or less horizontal line ignoring an increase in the lower and middle dose groups. Another situation with different outcomes is the

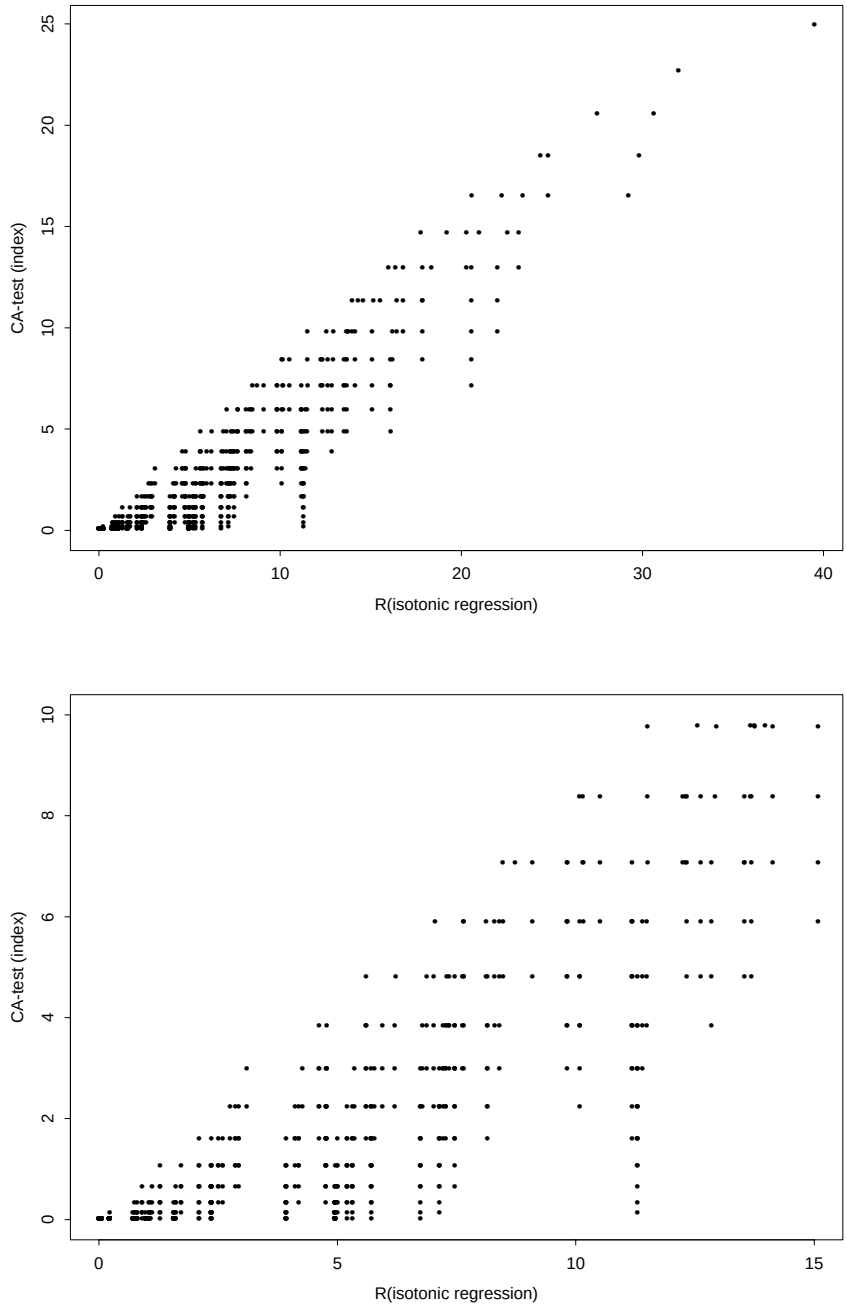

Figure 5: Comparison of results of isotonic regression (R) and the CA–test using the index as scores (χ^2_{slope}) for all 1001 possible combinations of the para–aramid study (see table 1). The upper graph contains all data, the lower one is restricted to $R \leq 15$ and $\chi^2_{slope} \leq 10$.

permutation $(4, 0, 0, 0, 6)$. Isotonic regression gives a test-statistic of $R = 11.18$, whereas the CA–test yields the following results: using the index $\chi^2_{slope} = 1.57$, using the dose $\chi^2_{slope} = 13.22$ and using log dose $\chi^2_{slope} = 0$.

To summarize the results obtained in connection with the para–aramid example the CA–test seems highly vulnerable towards deviations from linearity. It seems that the use of the dose as scores is not a good idea especially when the dose of the highest group is disproportionately larger.

6. DISCUSSION

Within this paper the analysis of a dose–response reationship between discrete levels of the dose and a binary response was considered. The commonly used Cochran–Armitage test was compared with the isotonic regression. The result of the CA–test depends highly on the form of the relationship as in linear regression analysis. The transformation of the x–axis in the CA–test using different score assignments can lead to different conclusions. There are situations in which one test rejected the null hypothesis and another one accepts it.

In order to solve this problem another test base on isotonic regression was proposed. The only assumption for this approach to work, is the monotonicity of the response. But this assumption is also required for many other tests, the CA–test included. No additional assumptions are made about the form of the relationship. Any monotonic transformation of the x–axis, in our example the dose–levels, leads to identitical results. The power of this approach is close to that of the optimal CA–test. There is only a slight difference in power of about 2.5% depending on the situation considered.

In order to obtain the estimates of the true p–values we recommend using permutation-tests. There are two options. Permutations can be generated on a random base. About 10000 replications seem sufficient to give p–values of adequate resolution. The other option is to consider all possible permutations, calculate the probability to observe this permutation and to add up all the probabilities for combinations with an equal or larger test–statistic.

The advantages of the permutation–test approach are twofold. The tabulated values of the distribution of the test–statistics are valid only for large samples and for equal number of observations per group. Both criteria are rarely fulfilled in reality. Applying isotonic regression to the data of the para–aramid study, the hypothesis of a dose– response relationship cannot be accepted. The conflict between the various results of the CA–test depending on the score assignments can be solved by this approach. Therefore we recommend performing the analysis of a dose–response relationship using isotonic regression.

ACKNOWLEDGEMENTS

This research was supported by grant UL 94/8-3 of the Deutsche Forschungsgemeinschaft.

REFERENCES

1. Neumann, H. 'Proposed changes in the classification of carciogenic chemicals in the work area', *Regulatory Toxicology and Pharmacology*, **26**, 288–295 (1997).
2. Chuang-Stein, C. and Agresti, A. 'Tutorial in biostatistics: A review of tests for detecting a monotone dose–response relationship with ordinal response data', *Statistics in Medicine*, **16**, 2599–2618 (1997).
3. Cochran, W. G. 'Some methods for strengthening common χ^2 tests', *Biometrics*, **10**, 417–451 (1954).
4. Armitage, P. 'Tests for linear trends in proportions and frequencies', *Biometrics*, **11**, 375–386 (1955).
5. IARC-Monographs . *Silica, Some Silicates, Coal Dust and para–Aramid fibrils*, World Health Organization, International Agency for Research on Cancer, 1997.
6. Robertson, T., Wright, F. T. and Dykstra, R. L. *Order Restricted Statistical Inference*, Wiley, New York, USA, 1988.
7. Fleiss, J. L. *Statistical Methods for Rates and Proportions*, J. Wiley, New York, 1981.
8. Chen, J. J., Kodell, R. and Pearce, B. 'Significance levels of randomization trend tests in the event of rare occurrences', *Biometrical Journal*, **39**, 327–337 (1997).