METHODOLOGY

# Analysis of graded lesions in long-term carcinogenicity studies

John S Fry
Peter N Lee
Jan S Hamling

PN Lee Statistics and Computing,
Sutton, UK

**Abstract:** In long-term carcinogenicity studies, the Peto test is the standard method for analyzing lesions that are not observable in life, taking into account differential between-group mortality. This method requires knowledge for each animal in each group of when the animal died, whether a lesion was seen postmortem, and the context of observation (ie, whether the lesion was fatal [considered to have contributed to its death] or whether it was incidental [seen in an animal dying for an unrelated reason]). The Peto test was not designed to take severity of the lesion into account. Age-adjusted analysis taking severity into account can be carried out using a stratified version of the Kruskal–Wallis one-way analysis of variance by ranks, but this does not take context of observation into account. Here, we describe an extended version of the stratified Kruskal–Wallis test that both allows analysis of lesions graded on severity and takes fatality of the lesion into account. This includes both a test for between-group heterogeneity and a test for dose-related trend, and provides a powerful method for analyzing such data. The Peto and Kruskal–Wallis tests may be considered special cases of this more general test. The test may be applied not only to nonneoplastic lesions (such as chronic progressive nephropathy), where pathologists often grade severity on a 5-point scale, but also to preneoplastic and neoplastic lesions, carrying out a single analysis of focal hyperplasia, benign tumor, and malignant tumor of a specific histological type, regarding these as successive stages of a progressive condition.

**Keywords:** carcinogenicity tests, survival analysis, heterogeneity, trend, stratified analysis

## Introduction

Analysis of the incidence of tumors in long-term experimental studies taking into account differential mortality between groups is not straightforward. The problem arises mainly because, with some exceptions (eg, skin tumors), the presence of the tumor is not known until postmortem, and the time of onset is not known at all. The standard method for analyzing tumors that are not observable in life is the Peto test.[1] In this method, the context of observation of a tumor seen in a decedent animal is critical. Tumors that directly or indirectly kill their hosts are said to be observed in a fatal context, while tumors that do not and that are merely seen at the autopsy of an animal that has died of some unrelated cause are said to be observed in an incidental context. The definition of context of observation and the underlying assumptions behind the Peto test are discussed in detail in the original article describing the test and in a later International Agency for Research on Cancer monograph.[1,2]

While the method is, in principle, also applicable to nonneoplastic lesions, it is only designed to deal with binary data, where the animal may or may not have the lesion of interest. Suppose, for example, we wish to analyze the incidence of chronic progressive

Correspondence: John S Fry
PN Lee Statistics and Computing,
Sutton, Surrey SM2 5DA, UK
Tel +44 208 642 8265
Fax +44 208 642 2135
Email johnfry@pnlee.co.uk

**II**

nephropathy (CPN), for which the pathologist has recorded severity on a scale of 0 = absent, 1 = minimal, 2 = slight, 3 = moderate, 4 = marked, and 5 = severe. It is clearly possible to apply the Peto test to a variable relating to presence or absence of the condition, providing the pathologist has recorded whether the condition contributed to the death of the animal or not. However, the effect of treatment might be to increase the severity of the condition rather than to increase the incidence of the condition, and such an approach would miss this. An alternative is to create variables using different cutoff values, such as "CPN at least moderate" or "CPN at least marked." However, not only does this cause problems relating to multiple testing (especially if analyses are run for all cutoffs [severity grades 1+, 2+, 3+, 4+, or 5]) but there is also a loss of power, in that none of these analyses considers the complete spectrum of the grades seen.

Analysis of incidental findings without age adjustment can readily be carried out by Kruskal–Wallis one-way analysis of variance by ranks,[3] for which a test of trend is also available.[4,5] In 1988, we described[6] an extension to these tests that would allow for stratification by other variables. Using grouped time period as the stratifying variable, and ranking findings by severity, this can be used to conduct age-adjusted analysis of incidental findings.

In this paper, we introduce a method that combines the two approaches, taking into account the context of observation of the lesion, the severity grade of the lesion, and the time of death of the animal, thus providing a powerful analysis for this type of data. We refer to this test as the Fry–Lee–Hamling test. Apart from analyzing nonneoplastic conditions, such as CPN, the method may also be used to carry out a single analysis of focal hyperplasia, benign tumor, and malignant tumor of a specific histological type, regarding the conditions as successive stages of a progressive condition.

In order to help the reader clearly understand our new test, we first summarize the Peto test and the stratified Kruskal–Wallis test, before describing the new test.

## The Peto test

The Peto test requires knowledge of time of death for each animal, whether the lesion of interest was present or not, and whether it was classified as fatal or incidental.

The method starts by analyzing the fatal lesions, dividing the time scale into small periods, typically of a week. In each group ($j = 1, \dots k$), and for each period considered in which a fatal lesion occurs, the number at risk is the number of animals alive and lesion-free at the start of the period, and the expected number of animals and the variances of these

expected numbers are calculated assuming that the lesions seen in the period were equally likely to have occurred in any of the groups under consideration. Thus, if the data for a period are as follows:

|  | Group j | Total |
|---|---|---|
| Number observed with lesion in the period | $a_j$ | $n_1$ |
| Number without lesion in the period | $c_j$ | $n_0$ |
| Total number at risk | $m_j$ | $N$ |

we have:

$e_j$, the expected value of $a_j = m_j n_1 / N$

$V_{jj}$, the variance of $a_j = m_j(N - m_j)n_1 n_0 / N^2(N - 1)$

$V_{ij}$, the covariance of $a_i$ and $a_j = -m_i m_j n_1 n_0 / N^2(N - 1)$, $i \neq j$

The observed and expected numbers and the variance and covariance terms are then summed over all time periods where fatal lesions occur. If $\underline{a}.$ is the vector of the first $k - 1$ observed values summed over all time periods, $\underline{e}.$ is the vector of the first $k - 1$ summed expected values, and $V.$ is the $(k - 1) \times (k - 1)$ summed covariance matrix, $H = (\underline{a}. - \underline{e}.)^T V.^{-1} (\underline{a}. - \underline{e}.)$ is then approximately distributed as chi-squared on $k - 1$ degrees of freedom (df). H tests for the heterogeneity of fatal lesion incidence over the groups.

The analysis of the incidental lesions is then conducted. The method is essentially similar, except for three major differences. First, any animal that ever had a fatal occurrence of the lesion does not contribute at all. Second, the at-risk population is the total number dying in the time period, so that prevalence rather than incidence is considered. Third, the time periods used have to be broad enough to provide an adequate number of animals to analyze, though still narrow enough to allow reasonable age adjustment. Given these differences, the formulae are identical, providing a test for heterogeneity of incidental lesion prevalence over group.

Finally, an overall test of heterogeneity of lesion occurrence, allowing for context of observation and time of death, can be obtained based on the observed and expected values, and their variances and covariances, summed over the fatal and incidental analyses combined. This test statistic, $H_{PETO}$, is the test that is of most interest, tests of heterogeneity of fatal and incidental lesions separately not usually being reported.

There are a number of points to note:
a. Setting the time periods for incidental analyses is not straightforward. The source paper[1] suggested a maximum-likelihood approach applied separately for each lesion. In practice, this was found not always to converge.

We favor an approach where the periods are set independently of the lesions and are based only on the survival distribution of the animals. A possible algorithm is given in Supplementary materials, Calculation of grouped periods of death for the incidental part of the Peto test. Alternatively, fixed intervals may be used, eg, 1–52, 53–78, 79–91, and 92–104 weeks for a 2-year study. Note that terminal (and interim) kills should generally be treated as separate time periods. However, sometimes all the animals in one group have a terminal kill at an earlier time than the animals in other groups. While this is not an ideal experimental design, it may be preferable to include these animals along with animals that died in the time period including the time of kill, so avoiding the loss of information that would occur otherwise.

b. The method can also be used to conduct pair-wise tests, by simply selecting the two groups of interest and running the same analysis.

c. The Peto method also includes a test for dose-related trend. If $d_j$ is the dose (or nominal dose) applied to dose group j, where i references time period (i = 1, … L), the trend T is defined as:

$$T = \sum_{j=1}^{k} d_j(a_j. - e_j.)$$

and its variance, VarT, is given by:

$$VarT = \sum_{i=1}^{L} n_{i0} \left( \sum_{j=1}^{k} \left( d_j^2 e_{ij} \right) - \left( \sum_{j=1}^{k} d_j e_{ij} \right)^2 \Big/ n_{i1} \right) \Big/ (N_i - 1)$$

where $n_{i0}$ is the number of subjects without the lesion in period i, and $n_{i1}$ is the number of subjects with the lesion in period i. $Z_{PETO} = T/\sqrt{VarT}$ is then approximately normally distributed and is the test statistic for dose-related trend.

d. Exact versions of the pairwise and trend tests can be calculated by treating each subtable produced (ie, the tables of occurrences by group for each time period for both the contexts of observation) as independent, assigning a score (usually proportional to the dose level) to having the lesion present, then calculating the probability of achieving greater than or equal to the observed score by convoluting all the scores and probabilities across the subtables. This is of particular use where there are very few occurrences, and the asymptotic probability tends to overestimate the significance of any differences observed. Note, however, that the assumptions of this exact test have been questioned.[7] The calculation of the exact trend test is

quite complex, but can be carried out using the counting algorithm given in Supplementary materials, The exact trend test.

e. The Peto test (and indeed all the tests described in this paper) can readily be stratified for variables other than time of death, such as sex, or batch (where randomization is conducted separately within batch) to provide an overall test adjusted for these variables. Here, the data are divided into separate strata, in each of which observed and expected values and their variances and covariances are estimated as described above, and then summed over strata to allow calculation of the overall heterogeneity and trend tests.

## The unstratified and stratified Kruskal–Wallis test

The analyses described in this section are appropriate for analyzing lesions that are graded by severity, without taking context of observation into account. This section covers unadjusted analyses (the Kruskal-Wallis test), and adjusted (stratified) analyses (the Fry-Lee test).

## The Kruskal–Wallis test

For the standard Kruskal–Wallis test, the relevant statistic is:

$$H_{KW} = \left[ 12 \sum_{j=1}^{k} \left( R_j^2/N_j \right) \Big/ \left( N(N+1) \right) - 3(N+1) \right] \Big/ C$$

where N is the total number of animals, k is the number of groups, $N_j$ is the number of animals in group j, $R_j$ is the sum of ranks of animals in group j, and C is the correction for ties, calculated as:

$$C = 1 - \left( \sum t^3 - \sum t \right) \Big/ (N^3 - N)$$

where t is the number of tied observations for a given tied rank. The analysis is ranked on severity, so subjects in all groups that have the same severity will have the same tied rank.

$H_{KW}$ is asymptotically distributed as chi-squared on k − 1 df and tests for heterogeneity between groups. This test statistic can be re-written, without needing to consider tied ranks separately, as:

$$H_{KW} = \sum_{j=1}^{k} \left( O_j - E. \right)^2 N_j/V$$

where $O_j$ is the mean rank for group j, E. is the expected rank assuming no difference between groups, and V is the

sample variance of the ranks assuming no difference between the groups. Defining $R_{ij}$ as the rank for animal i in group j, we have:

$$O_j = \sum_{i=1}^{N_j} R_{ij}/N_j,$$

$$E. = (N+1)/2, \quad \text{and}$$

$$V = \left( \sum_{j=1}^{k} \sum_{i=1}^{N_j} R_{ij}^2 - NE.^2 \right) \Big/ (N-1)$$

$H_{KW}$ expressed in this way is also distributed asymptotically as chi-squared on k − 1 df.

Presumably, the more complicated version was created when calculations had to be done by hand. A proof of the equivalence of these two formulations is given in Supplementary materials, Equivalence of the test statistics used in calculating Kruskal–Wallis statistics.

Where only two groups are considered, the Kruskal–Wallis analysis is equivalent to the Wilcoxon test (and the Mann–Whitney *U* test). Using the standard Kruskal–Wallis equations, it is possible to generate a trend statistic:

$$Z_{KW} = T/\sqrt{(C \operatorname{Var} T)}$$

where:

$$T = \sum_{j=1}^{k} d_j N_j D_j,$$

$$\operatorname{var} T = (N+1) \left( N \sum_{j=1}^{k} d_j^2 N_j - \left( \sum_{j=1}^{k} d_j N_j \right)^2 \right) \Big/ 12$$

$d_j$ is the dose level applied in group j, and $D_j = O_j - E.$ is the difference between the mean rank for group j and the overall mean rank.

$Z_{KW}$ is approximately normally distributed. Based on the result in Supplementary materials, Equivalence of the test statistics used in calculating Kruskal–Wallis statistics, this can be rewritten as:

$$Z_{KW} = T/\sqrt{(VW)}$$

where:

$$W = \sum_{j=1}^{k} d_j^2 N_j - \left( \sum_{j=1}^{k} d_j N_j \right)^2 \Big/ N$$

## The stratified Kruskal–Wallis test (the Fry–Lee test)

For a carcinogenicity study, it would be beneficial if the analyses were stratified by age and sex. Age stratification can be achieved using grouped time periods. As we described earlier,[6,8] a stratified version of the Kruskal–Wallis test can be derived by first scaling the observed mean sum of ranks and expected mean sum of ranks in each stratum by the factor Ns/2, where Ns is the number of animals in the stratum. Then, using a procedure similar to the Peto test, we sum the scaled mean ranks and also the appropriately scaled variances and covariances over all strata.

Thus, for stratum s and group j, we have:

$$Os_j = 2Rs_j/Ns,$$

$$Es_j = Ns_j(Ns+1)/Ns,$$

$$\operatorname{Var}(Os_j) = 4 Vs\, Ns_j(Ns - Ns_j)/Ns^3,$$

$$\operatorname{Cov}(Os_i, Os_j) = -4 Vs\, Ns_i Ns_j/Ns^3$$

where, for a given stratum, $Rs_j$ is the sum of ranks for group j, Ns is the total number of animals, and Vs is the sample variance of the ranks assuming no difference between the groups.

Given that $\underline{O}.$ is the vector formed by the sum of $Os_j$ for the first k − 1 groups over all the strata, $\underline{E}.$ is the corresponding vector derived from the $Es_j$, and V. is the matrix formed from the sum of the variance/covariance values for the same k − 1 groups over all the strata, the stratified Kruskal–Wallis statistic is:

$$H_{FL} = (\underline{O}. - \underline{E}.)^T V.^{-1} (\underline{O}. - \underline{E}.)$$

which is then approximately distributed as chi-squared on k − 1 df.

For the trend statistic, we use:

$$Z_{FL} = \sum_s Ts \Big/ \sqrt{\left( \sum_s UsWs \right)}$$

where:

$$Ts = \sum_{j=1}^{k} d_j \left( Os_j - Es_j \right),$$

$Us = 4Vs/Ns^2$, ie, the variance, appropriately scaled, and

$$Ws = \sum_{j=1}^{k} d_j^2 Ns_j - \left( \sum_{j=1}^{k} d_j Ns_j \right)^2 \Big/ Ns$$

$Z_{FL}$ is again approximately normally distributed.

With this scaling, the results for data with only two possible values are identical to the results for the stratified chi-squared test. The stratified chi-squared test can therefore be regarded simply as a special case of the stratified Kruskal–Wallis test.

Often a continuity correction is made to chi-squared statistics to improve the asymptotic properties. Known as Yates's correction, this can also be applied to pair-wise statistics, once the scaling has been allowed for. Thus, we replace $Os_j - Es_j$ by $(|Os_j - Es_j| - 0.5)$ to continuity-correct the values.

It is interesting to note that where we are performing pair-wise tests within a single stratum, the exact probabilities can be derived by restating the problem. While for the actual data, observations may be in either the control or the treated group, and may take severity values of 0–5 (assuming the graded lesion has five levels), we effectively reverse the rows and columns so that observations in the control group take the value "absent," observations in the treated group take the value "present," and the severity levels of the graded lesion become groups. Thus, starting with a table such as:

| Severity | Control | Treatment j |
|---|---|---|
| 0 | $a_{00}$ | $a_{j0}$ |
| 1 | $a_{01}$ | $a_{j1}$ |
| 2 | $a_{02}$ | $a_{j2}$ |
| 3 | $a_{03}$ | $a_{j3}$ |
| 4 | $a_{04}$ | $a_{j4}$ |
| 5 | $a_{05}$ | $a_{j5}$ |

we consider instead the table:

| | Group | | | | | |
|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | 5 |
| Present | $a_{j0}$ | $a_{j1}$ | $a_{j2}$ | $a_{j3}$ | $a_{j4}$ | $a_{j5}$ |
| Absent | $a_{00}$ | $a_{01}$ | $a_{02}$ | $a_{03}$ | $a_{04}$ | $a_{05}$ |

The problem has now become identical to that already solved for the Fisher exact trend test, and so the algorithm given in Supplementary materials, The exact trend test can be used.

## Extending the stratified Kruskal–Wallis test to cope with fatal and incidental lesions (the Fry–Lee–Hamling test)

Although the stratified Kruskal–Wallis test allows analyses to take account of age, sex, and the severity of the lesion, it does not take account of the context of observation of the lesion ("fatal" or "incidental").

There seems no inherent reason why the approaches of the Peto test and the stratified Kruskal–Wallis test should not be combined to produce an age-adjusted test that simultaneously adjusts for severity of a lesion, and whether the lesion was fatal or incidental. To do this involves two passes through the data. On the fatal pass, we take as the first time period the period with the first occurrence of the lesion presumed to have contributed to death. We then analyze the data on the animals at risk at the start of that time period, using the formulation from the stratified Kruskal–Wallis test to give us the scaled observed and expected vectors for the different groups and the appropriate variance/covariance matrix. We then continue to the next time point, adding up the vectors and matrices as we go. Then we remove from consideration any animals that had a fatal version of the lesion, and conduct a further set of stratified Kruskal–Wallis tests for the broader time periods where incidental occurrences of the lesion were seen, adding further to the accumulated vectors and matrices. Finally we use the formulae given above (as $H_{FL}$ and $Z_{FL}$) to calculate our test statistics, $H_{FLH}$ testing for heterogeneity on $k-1$ df and $Z_{FLH}$ testing for trend on 1 df.

## An example

In this section, we work through an example, using data from a real study including four groups of 50 male animals dosed at 0 (control), 10, 25, and 50 mg/kg/day. The lesion of interest is CPN. In the following sections, we first summarize the data set and the preliminary impressions to be gained from it. We then summarize the results of Peto analyses of CPN of any severity and of occurrences above specified severities, before giving the results and intermediate calculations relat-

**Table I** Frequency Table

| | Group | | | | Total |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | |
| *Dose (mg/kg/day)* | 0 | 10 | 25 | 50 | |
| **Kidney: chronic progressive nephropathy** | | | | | |
| Incidental | | | | | |
| ~none | 33 | 26 | 31 | 27 | 117 |
| ~minimal | 12 | 17 | 10 | 4 | 43 |
| ~slight | 2 | 4 | 3 | 2 | 11 |
| ~moderate | 1 | 1 | 3 | 2 | 7 |
| ~marked | 0 | 0 | 0 | 4 | 4 |
| ~severe | 0 | 0 | 0 | 5 | 5 |
| Fatal | | | | | |
| ~moderate | 1 | 0 | 0 | 0 | 1 |
| ~marked | 0 | 0 | 2 | 2 | 4 |
| ~severe | 1 | 2 | 1 | 4 | 8 |

**Table 2** Fisher's exact text

| | Group | | | | | Trend |
|---|---|---|---|---|---|---|
| | **1** | **2** | **3** | **4** | | |
| *Dose (mg/kg/day)* | **0** | **10** | **25** | **50** | | |
| **Kidney: chronic progressive nephropathy** | | | | | | |
| ~severe | | | | | | |
| n | 1 | 2 | 1 | 9 | | |
| N | 50 | 50 | 50 | 50 | | |
| % | 2.00 | 4.00 | 2.00 | 18.00 | | |
| P | | 1.0000 | 1.0000 | 0.0157+ | | 0.0015++ |
| ~severe or ~marked | | | | | | |
| n | 1 | 2 | 3 | 15 | | |
| N | 50 | 50 | 50 | 50 | | |
| % | 2.00 | 4.00 | 6.00 | 30.00 | | |
| P | | 1.0000 | 0.6173 | 0.0002+++ | | 0.0000+++ |
| At least ~moderate | | | | | | |
| n | 3 | 3 | 6 | 17 | | |
| N | 50 | 50 | 50 | 50 | | |
| % | 6.00 | 6.00 | 12.00 | 34.00 | | |
| P | | 1.0000 | 0.4870 | 0.0008+++ | | 0.0000+++ |
| At least ~slight | | | | | | |
| n | 5 | 7 | 9 | 19 | | |
| N | 50 | 50 | 50 | 50 | | |
| % | 10.00 | 14.00 | 18.00 | 38.00 | | |
| P | | 0.7596 | 0.3881 | 0.0019++ | | 0.0003+++ |
| Any grade | | | | | | |
| n | 17 | 24 | 19 | 23 | | |
| N | 50 | 50 | 50 | 50 | | |
| % | 34.00 | 48.00 | 38.00 | 46.00 | | |
| P | | 0.2223 | 0.8352 | 0.3074 | | 0.4532 |

**Abbreviations:** +++, $P < 0.001$; ++, $P < 0.01$; +, $P < 0.05$; (+), $P < 0.1$.

ing to the Fry–Lee–Hamling tests for heterogeneity and trend. Associated with this example are two sets of data. Numbered tables give more detailed output corresponding to the example in this section, and also give results from the poly-$k$ test referred to in the Conclusions section. These tables also give the results for pair-wise tests versus the control group, and provide different sets of results according to whether or not Yates's correction is used. The other set of data is an Example analyses file that gives the raw data and works through the example in this section, giving the Peto analysis and the Fry–Lee–Hamling analysis in full. These data also give examples of the exact test, including exact trend analysis (as described in Supplementary materials, The exact trend test).

## The data and preliminary impressions

Table 1 shows the detailed frequencies by group, and Table 2 the results of applying Fisher's exact test to the data. The table below shows overall frequencies of fatal and incidental occurrences (n), without reference to severity, and numbers at risk (N). There is a slight excess of fatal occurrences in group 4 and of incidental occurrences in group 2, but no clear treatment effect is evident.

| | Group | | | |
|---|---|---|---|---|
| | **1** | **2** | **3** | **4** |
| Fatal | | | | |
| n | 2 | 2 | 3 | 6 |
| N | 50 | 50 | 50 | 50 |
| % | 4.0000 | 4.0000 | 6.0000 | 12.0000 |
| Incidental | | | | |
| n | 15 | 22 | 16 | 17 |
| N | 48 | 48 | 47 | 44 |
| % | 31.2500 | 45.8333 | 34.0426 | 38.6364 |
| Total | | | | |
| n | 17 | 24 | 19 | 23 |
| N | 50 | 50 | 50 | 50 |
| % | 34.0000 | 48.0000 | 38.0000 | 46.0000 |

Looking at the distribution of the levels of severity, we see a different picture, with a far greater frequency of the more severe forms of the lesion in the top dose – group 4.

| **Severity** | | Group | | | |
|---|---|---|---|---|---|
| | | **1** | **2** | **3** | **4** |
| Severe | n | 1 | 2 | 1 | 9 |
| Marked | n | 0 | 0 | 2 | 6 |
| Moderate | n | 2 | 1 | 3 | 2 |
| Slight | n | 2 | 4 | 3 | 2 |
| Minimal | n | 12 | 17 | 10 | 4 |

## Peto analyses

The standard Peto analyses of CPN of any severity, and of occurrences above specified severities, bear out this impression. In the table below, the probabilities for the pair-wise analyses compared with the control group (group 1) are shown in the column for the relevant dosed group, while the probabilities for the heterogeneity and trend tests are shown in the final two columns. The symbols +++, ++, +, and (+) are used to indicate positive pair-wise differences from the control group or trends that are significant at $P < 0.001$, <0.01, <0.05, or <0.1, with asterisks used for heterogeneity, which is nondirectional. Ex $P$ represents the Peto test probabilities calculated using exact methods, and Chi $P$ represents the Peto Test asymptotic probabilities (see Tables 3 and 4). The two sets of probabilities are very close in this instance, as the number of occurrences is reasonably large.

| Severity | Group | | | | Total | Trend |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | | |
| Severe | | | | | | |
| n | 1 | 2 | 1 | 9 | | |
| % | 2.00 | 4.00 | 2.00 | 18.00 | | |
| Ex $P$ | | 0.9823 | 1.0000 | 0.0385+ | | 0.0041++ |
| Chi $P$ | | 0.9837 | 0.4982 | 0.0460+ | 0.0084** | 0.0029++ |
| Severe or marked | | | | | | |
| n | 1 | 2 | 3 | 15 | | |
| % | 2.00 | 4.00 | 6.00 | 30.00 | | |
| Ex $P$ | | 0.9823 | 0.6711 | 0.0015++ | | 0.0000+++ |
| Chi $P$ | | 0.9837 | 0.6602 | 0.0029++ | 0.0001*** | 0.0000+++ |
| At least moderate | | | | | | |
| n | 3 | 3 | 6 | 17 | | |
| % | 6.00 | 6.00 | 12.00 | 34.00 | | |
| Ex $P$ | | 1.0000 | 0.5645 | 0.0051++ | | 0.0001+++ |
| Chi $P$ | | 0.7745 | 0.5605 | 0.0066++ | 0.0004*** | 0.0000+++ |
| At least slight | | | | | | |
| n | 5 | 7 | 9 | 19 | | |
| % | 10.00 | 14.00 | 18.00 | 38.00 | | |
| Ex $P$ | | 0.6437 | 0.4153 | 0.0043++ | | 0.0004+++ |
| Chi $P$ | | 0.6446 | 0.4154 | 0.0052++ | 0.0028** | 0.0003+++ |
| Any grade | | | | | | |
| n | 17 | 24 | 19 | 23 | | |
| % | 34.00 | 48.00 | 38.00 | 46.00 | | |
| Ex $P$ | | 0.0499+ | 0.8688 | 0.0876(+) | | 0.0733(+) |
| Chi $P$ | | 0.0534(+) | 0.8702 | 0.0904(+) | 0.0486* | 0.0691(+) |

In this example, the size of the effect makes it fairly clear that there really is a treatment effect present, but only at higher severity levels. But in order to show this, we have had to perform five sets of analyses. With smaller effects, it would be difficult to discriminate real effects and effects that only appear significant due to multiple testing. We now try our combined analysis.

## Using the Fry–Lee–Hamling test for heterogeneity

First we carry out the fatal analysis, analyzing each week in which a fatal occurrence was seen. The first fatal occurrence was seen at week 56 in group 4. Ten animals had died prior to that week, giving the table shown below.

Fatal lesions in week 56

| | Group | | | | Total |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | |
| Severity 0 | 48 | 46 | 49 | 47 | 190 |
| 1–4 | 0 | 0 | 0 | 0 | 0 |
| 5 | 0 | 0 | 0 | 1 | 1 |
| Total ($N_j$) | 48 | 46 | 49 | 48 | 191 |

We have 190 animals with no lesion (grade 0), each having a mid-rank of $(1 + 190)/2 = 95.5$. The final animal has rank 191. The expected rank (E.) is given by $(N + 1)/2 = 96$. We can then calculate:

Fatal lesions in week 56

| | Group | | | | Total |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | |
| Sum of ranks for severity 0 | 4584 | 4393 | 4679.5 | 4488.5 | 18145 |
| Sum of ranks for severity 5 | 0 | 0 | 0 | 191 | 191 |
| Observed mean rank ($O_j$) | 95.5 | 95.5 | 95.5 | 97.48958 | 96 |
| $D_j = O_j - E.$ | −0.5 | −0.5 | −0.5 | 1.489583 | |
| $D_j N_j$ | −24 | −23 | −24.5 | 71.5 | |
| $D_j^2 N_j$ | 12 | 11.5 | 12.25 | 106.5052 | 142.2552 |
| Sum of squares of ranks for severity 0 | 437772 | 419531.5 | 446892.3 | 428651.8 | |
| Sum of squares of ranks for severity 5 | 0 | 0 | 0 | 36481 | |
| Sum of squares of ranks | 437772 | 419531.5 | 446892.3 | 465132.8 | 1769329.5 |

We then use these results and the formulae given earlier to calculate the following:

Fatal lesions in week 56

| | Group | | | | Total |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | |
| Variance/ covariance matrix | 0.18815 | −0.06052 | −0.06447 | −0.06316 | |
| | −0.06052 | 0.18283 | −0.06179 | −0.06052 | |
| | −0.06447 | −0.06179 | 0.19073 | −0.06447 | |
| | −0.06316 | −0.06052 | −0.06447 | 0.18815 | |
| Inverse of matrix (groups 1–3) | 7.958333 | 3.979167 | 3.979167 | | |
| | 3.979167 | 8.131341 | 3.979167 | | |
| | 3.979167 | 3.979167 | 7.877126 | | |
| $(Os_j - Es_j)$, scaled | −0.25131 | −0.24084 | −0.25654 | 0.74869 | 0 |

**Table 3** Peto test (exact probabliities)

| | Group | | | | Trend |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | |
| *Dose (mg/kg/day)* | 0 | 10 | 25 | 50 | |
| **Kidney: chronic progressive nephropathy** | | | | | |
| ~severe | | | | | |
|   Fatal | | | | | |
|     n | 1 | 2 | 1 | 4 | |
|     N | 50 | 50 | 50 | 50 | |
|     % | 2.00 | 4.00 | 2.00 | 8.00 | |
|     P | | 0.9823 | 1.0000 | 0.3454 | 0.1775 |
|   Incidental | | | | | |
|     n | 0 | 0 | 0 | 5 | |
|     N | 49 | 48 | 49 | 46 | |
|     % | 0.00 | 0.00 | 0.00 | 10.87 | |
|     P | | NS | NS | 0.1347 | 0.0062[++] |
|   Total | | | | | |
|     n | 1 | 2 | 1 | 9 | |
|     N | 50 | 50 | 50 | 50 | |
|     % | 2.00 | 4.00 | 2.00 | 18.00 | |
|     P | | 0.9823 | 1.0000 | 0.0385[+] | 0.0041[++] |
| ~severe or ~marked | | | | | |
|   Fatal | | | | | |
|     n | 1 | 2 | 3 | 6 | |
|     N | 50 | 50 | 50 | 50 | |
|     % | 2.00 | 4.00 | 6.00 | 12.00 | |
|     P | | 0.9823 | 0.6711 | 0.1100 | 0.0380[+] |
|   Incidental | | | | | |
|     n | 0 | 0 | 0 | 9 | |
|     N | 49 | 48 | 47 | 44 | |
|     % | 0.00 | 0.00 | 0.00 | 20.45 | |
|     P | | NS | NS | 0.0143[+] | 0.0000[+++] |
|   Total | | | | | |
|     n | 1 | 2 | 3 | 15 | |
|     N | 50 | 50 | 50 | 50 | |
|     % | 2.00 | 4.00 | 6.00 | 30.00 | |
|     P | | 0.9823 | 0.6711 | 0.0015[++] | 0.0000[+++] |
| At least ~moderate | | | | | |
|   Fatal | | | | | |
|     n | 2 | 2 | 3 | 6 | |
|     N | 50 | 50 | 50 | 50 | |
|     % | 4.00 | 4.00 | 6.00 | 12.00 | |
|     P | | 1.0000 | 1.0000 | 0.2385 | 0.0838[(+)] |
|   Incidental | | | | | |
|     n | 1 | 1 | 3 | 11 | |
|     N | 48 | 48 | 47 | 44 | |
|     % | 2.08 | 2.08 | 6.38 | 25.00 | |
|     P | | 1.0000 | 0.6612 | 0.0163[+] | 0.0002[+++] |
|   Total | | | | | |
|     n | 3 | 3 | 6 | 17 | |
|     N | 50 | 50 | 50 | 50 | |
|     % | 6.00 | 6.00 | 12.00 | 34.00 | |
|     P | | 1.0000 | 0.5645 | 0.0051[++] | 0.0001[+++] |
| At least ~slight | | | | | |
|   Fatal | | | | | |
|     n | 2 | 2 | 3 | 6 | |
|     N | 50 | 50 | 50 | 50 | |
|     % | 4.00 | 4.00 | 6.00 | 12.00 | |
|     P | | 1.0000 | 1.0000 | 0.2385 | 0.0838[(+)] |
|   Incidental | | | | | |
|     n | 3 | 5 | 6 | 13 | |
|     N | 48 | 48 | 47 | 44 | |

*(Continued)*

**Table 3** (*Continued*)

| | Group | | | | Trend |
| | 1 | 2 | 3 | 4 | |
|---|---|---|---|---|---|
| % | 6.25 | 10.42 | 12.77 | 29.55 | |
| P | | 0.5954 | 0.4595 | 0.0157+ | 0.0021++ |
| Total | | | | | |
| n | 5 | 7 | 9 | 19 | |
| N | 50 | 50 | 50 | 50 | |
| % | 10.00 | 14.00 | 18.00 | 38.00 | |
| P | | 0.6437 | 0.4153 | 0.0043++ | 0.0004+++ |
| Any grade | | | | | |
| Fatal | | | | | |
| n | 2 | 2 | 3 | 6 | |
| N | 50 | 50 | 50 | 50 | |
| % | 4.00 | 4.00 | 6.00 | 12.00 | |
| P | | 1.0000 | 1.0000 | 0.2385 | 0.0838(+) |
| Incidental | | | | | |
| n | 15 | 22 | 16 | 17 | |
| N | 48 | 48 | 47 | 44 | |
| % | 31.25 | 45.83 | 34.04 | 38.64 | |
| P | | 0.0288+ | 1.0000 | 0.3111 | 0.3245 |
| Total | | | | | |
| n | 17 | 24 | 19 | 23 | |
| N | 50 | 50 | 50 | 50 | |
| % | 34.00 | 48.00 | 38.00 | 46.00 | |
| P | | 0.0499+ | 0.8688 | 0.0876(+) | 0.0733(+) |

**Abbreviations:** +++, $P < 0.001$; ++, $P < 0.01$; +, $P < 0.05$; (+), $P < 0.1$.

Note that we can calculate the heterogeneity of this single time period using $H_{FL} = (Os_j - Es_j)^T V^{-1}(Os_j - Es_j)$, which gives a value of 2.979 on 3 df, with the corresponding probability 0.3948. This result is identical to that given by the Kruskal–Wallis test for this single time period (see Supplementary materials, Example analyses file).

We then apply the same technique to each of the other time periods, add up the variance/covariance matrices and the scaled differences $(Os_j - Es_j)$ over all the time periods, and calculate our test statistic for fatal lesions using the same formula for H.

All fatal lesions combined

| | Group | | | |
| | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Variance/ | 2.47030 | −0.80385 | −0.87864 | −0.78781 |
| covariance | −0.80385 | 2.38231 | −0.83057 | −0.74789 |
| matrix | −0.87864 | −0.83057 | 2.52414 | −0.81493 |
| | −0.78781 | −0.74789 | −0.81493 | 2.35064 |
| Inverse of | 0.625768 | 0.324296 | 0.324536 | |
| matrix | 0.324296 | 0.642218 | 0.324207 | |
| | 0.324536 | 0.324207 | 0.615824 | |
| $(Os_j - Es_j)$, scaled | −1.32338 | −1.14646 | −0.43594 | |

The test statistic for all fatal lesions combined is estimated as 3.740 on 3 df, with the corresponding probability 0.2910.

We now remove the fatal lesions and repeat the analysis using animals dying in particular grouped time periods and those killed terminally. Here, seven periods were used: the deaths in weeks 1–49, 50–76, 77–91, 92–97, 98–101, and 102–107, and the terminal killing in weeks 105–107, which usually includes most deaths. For that period, the data are:

Incidental lesions in terminal killing

| | Group | | | | Total |
| | 1 | 2 | 3 | 4 | |
|---|---|---|---|---|---|
| Severity | | | | | |
| 0 | 16 | 6 | 19 | 9 | 50 |
| 1 | 12 | 16 | 9 | 3 | 40 |
| 2 | 2 | 3 | 2 | 2 | 9 |
| 3 | 1 | 1 | 3 | 1 | 6 |
| 4 | 0 | 0 | 0 | 0 | 0 |
| 5 | 0 | 0 | 0 | 1 | 1 |
| Total ($N_j$) | 31 | 26 | 33 | 16 | 106 |

We have 50 animals at severity 0, each having a mid-rank of $(1 + 50)/2 = 25.5$. The 40 animals with severity 1 share ranks 51–90, thus each having a mid-rank of $(51 + 90)/2 = 70.5$. Similarly, the mid-ranks are 95 for severity 2, 102.5 for severity 3, and 106 for severity 5.

**Table 4** Peto test (asymptotic probabilities)

| | Group | | | | Total | Trend |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | | |
| *Yates's correction on* | | | | | | |
| **Dose (mg/kg/day)** | **0** | **10** | **25** | **50** | | |
| **Kidney: chronic progressive nephropathy** | | | | | | |
| ~severe | | | | | | |
|   Fatal | | | | | | |
|     n | 1 | 2 | 1 | 4 | 8 | |
|     E | 2.04 | 1.96 | 2.08 | 1.92 | 8.00 | |
|     E(2) | | 1.48 | 1.02 | 2.44 | | |
|     $\chi^2$ | | 0.00 | 0.46 | 0.90 | 3.35 | 2.15 |
|     P | | 0.9837 | 0.4982 | 0.3428 | 0.3405 | 0.1424 |
|   Incidental | | | | | | |
|     n | 0 | 0 | 0 | 5 | 5 | |
|     E | 1.00 | 1.25 | 0.98 | 1.78 | 5.00 | |
|     E(2) | | 0.00 | 0.00 | 3.09 | | |
|     $\chi^2$ | | 0.00 | 0.00 | 2.05 | 10.45 | 8.56 |
|     P | | NS | NS | 0.1522 | 0.0151* | 0.0034++ |
|   Total | | | | | | |
|     n | 1 | 2 | 1 | 9 | 13 | |
|     E | 3.04 | 3.20 | 3.06 | 3.70 | 13.00 | |
|     E(2) | | 1.48 | 1.02 | 5.53 | | |
|     $\chi^2$ | | 0.00 | 0.46 | 3.98 | 11.72 | 8.87 |
|     P | | 0.9837 | 0.4982 | 0.0460+ | 0.0084** | 0.0029++ |
| ~severe or ~marked | | | | | | |
|   Fatal | | | | | | |
|     n | 1 | 2 | 3 | 6 | 12 | |
|     E | 3.05 | 2.92 | 3.14 | 2.89 | 12.00 | |
|     E(2) | | 1.48 | 2.06 | 3.42 | | |
|     $\chi^2$ | | 0.00 | 0.19 | 2.48 | 5.01 | 4.87 |
|     P | | 0.9837 | 0.6602 | 0.1153 | 0.1708 | 0.0273+ |
|   Incidental | | | | | | |
|     n | 0 | 0 | 0 | 9 | 9 | |
|     E | 1.76 | 2.45 | 1.61 | 3.18 | 9.00 | |
|     E(2) | | 0.00 | 0.00 | 5.66 | | |
|     $\chi^2$ | | 0.00 | 0.00 | 5.15 | 20.24 | 16.78 |
|     P | | NS | NS | 0.0232+ | 0.0002*** | 0.0000+++ |
|   Total | | | | | | |
|     n | 1 | 2 | 3 | 15 | 21 | |
|     E | 4.80 | 5.37 | 4.75 | 6.07 | 21.00 | |
|     E(2) | | 1.48 | 2.06 | 9.09 | | |
|     $\chi^2$ | | 0.00 | 0.19 | 8.88 | 21.01 | 18.83 |
|     P | | 0.9837 | 0.6602 | 0.0029++ | 0.0001*** | 0.0000+++ |
| At least ~moderate | | | | | | |
|   Fatal | | | | | | |
|     n | 2 | 2 | 3 | 6 | 13 | |
|     E | 3.32 | 3.15 | 3.44 | 3.09 | 13.00 | |
|     E(2) | | 1.94 | 2.57 | 3.84 | | |
|     $\chi^2$ | | 0.19 | 0.00 | 1.39 | 3.74 | 3.36 |
|     P | | 0.6619 | 0.9471 | 0.2392 | 0.2910 | 0.0666(+) |
|   Incidental | | | | | | |
|     n | 1 | 1 | 3 | 11 | 16 | |
|     E | 3.64 | 4.19 | 3.55 | 4.62 | 16.00 | |
|     E(2) | | 0.91 | 2.06 | 7.14 | | |
|     $\chi^2$ | | 0.35 | 0.20 | 5.26 | 16.07 | 14.79 |
|     P | | 0.5548 | 0.6538 | 0.0218+ | 0.0011** | 0.0001+++ |

(*Continued*)

**20**

Open Access Medical Statistics 2013:3

**Table 4** (*Continued*)

| | Group | | | | Total | Trend |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | | |
| Total | | | | | | |
| n | 3 | 3 | 6 | 17 | 29 | |
| E | 6.97 | 7.34 | 6.98 | 7.71 | 29.00 | |
| E(2) | | 2.85 | 4.64 | 10.99 | | |
| $\chi^2$ | | 0.08 | 0.34 | 7.37 | 18.07 | 16.46 |
| P | | 0.7745 | 0.5605 | 0.0066[++] | 0.0004*** | 0.0000[+++] |
| At least ~slight | | | | | | |
| Fatal | | | | | | |
| n | 2 | 2 | 3 | 6 | 13 | |
| E | 3.32 | 3.15 | 3.44 | 3.09 | 13.00 | |
| E(2) | | 1.94 | 2.57 | 3.84 | | |
| $\chi^2$ | | 0.19 | 0.00 | 1.39 | 3.74 | 3.36 |
| P | | 0.6619 | 0.9471 | 0.2392 | 0.2910 | 0.0666[(+)] |
| Incidental | | | | | | |
| n | 3 | 5 | 6 | 13 | 27 | |
| E | 6.64 | 6.95 | 6.53 | 6.88 | 27.00 | |
| E(2) | | 3.79 | 4.46 | 8.50 | | |
| $\chi^2$ | | 0.28 | 0.54 | 5.57 | 10.68 | 9.96 |
| P | | 0.5972 | 0.4612 | 0.0182[+] | 0.0136* | 0.0016[++] |
| Total | | | | | | |
| n | 5 | 7 | 9 | 19 | 40 | |
| E | 9.96 | 10.09 | 9.97 | 9.98 | 40.00 | |
| E(2) | | 5.73 | 7.03 | 12.35 | | |
| $\chi^2$ | | 0.21 | 0.66 | 7.81 | 14.09 | 13.09 |
| P | | 0.6446 | 0.4154 | 0.0052[++] | 0.0028** | 0.0003[+++] |
| Any grade | | | | | | |
| Fatal | | | | | | |
| n | 2 | 2 | 3 | 6 | 13 | |
| E | 3.32 | 3.15 | 3.44 | 3.09 | 13.00 | |
| E(2) | | 1.94 | 2.57 | 3.84 | | |
| $\chi^2$ | | 0.19 | 0.00 | 1.39 | 3.74 | 3.36 |
| P | | 0.6619 | 0.9471 | 0.2392 | 0.2910 | 0.0666[(+)] |
| Incidental | | | | | | |
| n | 15 | 22 | 16 | 17 | 70 | |
| E | 19.07 | 17.49 | 19.52 | 13.92 | 70.00 | |
| E(2) | | 17.31 | 15.54 | 14.41 | | |
| $\chi^2$ | | 4.56 | 0.00 | 1.01 | 6.57 | 1.02 |
| P | | 0.0327[+] | 0.9862 | 0.3141 | 0.0871[(*)] | 0.3121 |
| Total | | | | | | |
| n | 17 | 24 | 19 | 23 | 83 | |
| E | 22.39 | 20.64 | 22.95 | 17.02 | 83.00 | |
| E(2) | | 19.25 | 18.11 | 18.25 | | |
| $\chi^2$ | | 3.73 | 0.03 | 2.87 | 7.88 | 3.30 |
| P | | 0.0534[(+)] | 0.8702 | 0.0904[(+)] | 0.0486* | 0.0691[(+)] |
| *Yates's correction off* | | | | | | |
| ~severe | | | | | | |
| Fatal | | | | | | |
| n | 1 | 2 | 1 | 4 | 8 | |
| E | 2.04 | 1.96 | 2.08 | 1.92 | 8.00 | |
| E(2) | | 1.48 | 1.02 | 2.44 | | |
| $\chi^2$ | | 0.36 | 0.00 | 1.95 | 3.35 | 2.15 |
| P | | 0.5499 | 0.9760 | 0.1625 | 0.3405 | 0.1424 |
| Incidental | | | | | | |
| n | 0 | 0 | 0 | 5 | 5 | |
| E | 1.00 | 1.25 | 0.98 | 1.78 | 5.00 | |

(*Continued*)

**Table 4** (*Continued*)

| | Group | | | | Total | Trend |
|---|---|---|---|---|---|---|
| | **1** | **2** | **3** | **4** | | |
| E(2) | | 0.00 | 0.00 | 3.09 | | |
| $\chi^2$ | | 0.00 | 0.00 | 3.76 | 10.45 | 8.56 |
| P | | NS | NS | 0.0526(+) | 0.0151* | 0.0034++ |
| **Total** | | | | | | |
| n | 1 | 2 | 1 | 9 | 13 | |
| E | 3.04 | 3.20 | 3.06 | 3.70 | 13.00 | |
| E(2) | | 1.48 | 1.02 | 5.53 | | |
| $\chi^2$ | | 0.36 | 0.00 | 5.43 | 11.72 | 8.87 |
| P | | 0.5499 | 0.9760 | 0.0198+ | 0.0084** | 0.0029++ |
| **~severe or ~marked** | | | | | | |
| **Fatal** | | | | | | |
| n | 1 | 2 | 3 | 6 | 12 | |
| E | 3.05 | 2.92 | 3.14 | 2.89 | 12.00 | |
| E(2) | | 1.48 | 2.06 | 3.42 | | |
| $\chi^2$ | | 0.36 | 0.88 | 3.82 | 5.01 | 4.87 |
| P | | 0.5499 | 0.3472 | 0.0507(+) | 0.1708 | 0.0273+ |
| **Incidental** | | | | | | |
| n | 0 | 0 | 0 | 9 | 9 | |
| E | 1.76 | 2.45 | 1.61 | 3.18 | 9.00 | |
| E(2) | | 0.00 | 0.00 | 5.66 | | |
| $\chi^2$ | | 0.00 | 0.00 | 7.13 | 20.24 | 16.78 |
| P | | NS | NS | 0.0076++ | 0.0002*** | 0.0000+++ |
| **Total** | | | | | | |
| n | 1 | 2 | 3 | 15 | 21 | |
| E | 4.80 | 5.37 | 4.75 | 6.07 | 21.00 | |
| E(2) | | 1.48 | 2.06 | 9.09 | | |
| $\chi^2$ | | 0.36 | 0.88 | 10.60 | 21.01 | 18.83 |
| P | | 0.5499 | 0.3472 | 0.0011++ | 0.0001*** | 0.0000+++ |
| **At least ~moderate** | | | | | | |
| **Fatal** | | | | | | |
| n | 2 | 2 | 3 | 6 | 13 | |
| E | 3.32 | 3.15 | 3.44 | 3.09 | 13.00 | |
| E(2) | | 1.94 | 2.57 | 3.84 | | |
| $\chi^2$ | | 0.00 | 0.15 | 2.35 | 3.74 | 3.36 |
| P | | 0.9496 | 0.7031 | 0.1255 | 0.2910 | 0.0666(+) |
| **Incidental** | | | | | | |
| n | 1 | 1 | 3 | 11 | 16 | |
| E | 3.64 | 4.19 | 3.55 | 4.62 | 16.00 | |
| E(2) | | 0.91 | 2.06 | 7.14 | | |
| $\chi^2$ | | 0.02 | 0.92 | 6.94 | 16.07 | 14.79 |
| P | | 0.9000 | 0.3365 | 0.0084++ | 0.0011** | 0.0001+++ |
| **Total** | | | | | | |
| n | 3 | 3 | 6 | 17 | 29 | |
| E | 6.97 | 7.34 | 6.98 | 7.71 | 29.00 | |
| E(2) | | 2.85 | 4.64 | 10.99 | | |
| $\chi^2$ | | 0.02 | 0.84 | 8.77 | 18.07 | 16.46 |
| P | | 0.9015 | 0.3580 | 0.0031++ | 0.0004*** | 0.0000+++ |
| **At least ~slight** | | | | | | |
| **Fatal** | | | | | | |
| n | 2 | 2 | 3 | 6 | 13 | |
| E | 3.32 | 3.15 | 3.44 | 3.09 | 13.00 | |
| E(2) | | 1.94 | 2.57 | 3.84 | | |
| $\chi^2$ | | 0.00 | 0.15 | 2.35 | 3.74 | 3.36 |
| P | | 0.9496 | 0.7031 | 0.1255 | 0.2910 | 0.0666(+) |

(*Continued*)

**22**

Open Access Medical Statistics 2013:3

**Table 4** (*Continued*)

| | Group | | | | Total | Trend |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | | |
| Incidental | | | | | | |
| n | 3 | 5 | 6 | 13 | 27 | |
| E | 6.64 | 6.95 | 6.53 | 6.88 | 27.00 | |
| E(2) | | 3.79 | 4.46 | 8.50 | | |
| $\chi^2$ | | 0.81 | 1.19 | 7.05 | 10.68 | 9.96 |
| P | | 0.3670 | 0.2755 | 0.0079++ | 0.0136* | 0.0016++ |
| Total | | | | | | |
| n | 5 | 7 | 9 | 19 | 40 | |
| E | 9.96 | 10.09 | 9.97 | 9.98 | 40.00 | |
| E(2) | | 5.73 | 7.03 | 12.35 | | |
| $\chi^2$ | | 0.58 | 1.19 | 9.13 | 14.09 | 13.09 |
| P | | 0.4469 | 0.2749 | 0.0025++ | 0.0028** | 0.0003+++ |
| Any grade | | | | | | |
| Fatal | | | | | | |
| n | 2 | 2 | 3 | 6 | 13 | |
| E | 3.32 | 3.15 | 3.44 | 3.09 | 13.00 | |
| E(2) | | 1.94 | 2.57 | 3.84 | | |
| $\chi^2$ | | 0.00 | 0.15 | 2.35 | 3.74 | 3.36 |
| P | | 0.9496 | 0.7031 | 0.1255 | 0.2910 | 0.0666(+) |
| Incidental | | | | | | |
| n | 15 | 22 | 16 | 17 | 70 | |
| E | 19.07 | 17.49 | 19.52 | 13.92 | 70.00 | |
| E(2) | | 17.31 | 15.54 | 14.41 | | |
| $\chi^2$ | | 5.72 | 0.05 | 1.56 | 6.57 | 1.02 |
| P | | 0.0168+ | 0.8258 | 0.2122 | 0.0871(*) | 0.3121 |
| Total | | | | | | |
| n | 17 | 24 | 19 | 23 | 83 | |
| E | 22.39 | 20.64 | 22.95 | 17.02 | 83.00 | |
| E(2) | | 19.25 | 18.11 | 18.25 | | |
| $\chi^2$ | | 4.66 | 0.14 | 3.58 | 7.88 | 3.30 |
| P | | 0.0308+ | 0.7091 | 0.0584(+) | 0.0486* | 0.0691(+) |

**Abbreviations:** E, expected value considering all groups; E(2), expected value considering only a single treated group and the control group; NS, not significant; +++ and ***, $P < 0.001$; ++ and **, $P < 0.01$; + and *, $P < 0.05$; (+) and (*), $P < 0.1$.

The expected rank (E.) is given by $(N + 1)/2 = 53.5$. We can calculate:

Incidental lesions in terminal killing

| | Group | | | | Total |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | |
| Sum of ranks for severity 0 | 408 | 153 | 484.5 | 229.5 | 1275 |
| Sum of ranks for severity 1 | 846 | 1128 | 634.5 | 211.5 | 2820 |
| Sum of ranks for severity 2 | 190 | 285 | 190 | 190 | 855 |
| Sum of ranks for severity 3 | 102.5 | 102.5 | 307.5 | 102.5 | 615 |
| Sum of ranks for severity 5 | 0 | 0 | 0 | 106 | 106 |
| Observed mean rank (O) | 49.88710 | 64.17308 | 48.98485 | 52.46875 | 53.5 |
| $D_j = O_j − E.$ | −3.6129 | 10.67308 | −4.51515 | −1.03125 | 1.51378 |

(*Continued*)

(*Continued*)

| | Group | | | | Total |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | |
| $D_j N_j$ | −112 | 277.5 | −149 | −16.5 | 0.0 |
| $D_j^2 \times N_j$ | 404.6452 | 2961.779 | 672.7576 | 17.01563 | 4056.1972 |
| Sum of squares of ranks for severity 0 | 10404 | 3901.5 | 12354.75 | 5852.25 | 32512.5 |
| Sum of squares of ranks for severity 1 | 59643 | 79524 | 44732.25 | 14910.75 | 198810 |
| Sum of squares of ranks for severity 2 | 18050 | 27075 | 18050 | 18050 | 81225 |
| Sum of squares of ranks for severity 3 | 10506.25 | 10506.25 | 31518.75 | 10506.25 | 63037.5 |
| Sum of squares of ranks for severity 5 | 0 | 0 | 0 | 11236 | 11236 |
| Sum of squares of ranks | 98603.25 | 121006.75 | 106655.75 | 60555.25 | 386821 |

We then use these results and the formulae given earlier to calculate the following:

**Incidental lesions in terminal killing**

| | Group | | | | Total |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | |
| Variance/ covariance matrix | 6.20382 | −2.15066 | −2.72968 | −1.32348 | |
| | −2.15066 | 5.55009 | −2.28941 | −1.11002 | |
| | −2.72968 | −2.28941 | 6.42796 | −1.40887 | |
| | −1.32348 | −1.11002 | −1.40887 | 3.844237 | |
| Inverse of matrix (groups 1–3) | 0.335023 | 0.220972 | 0.220972 | | |
| | 0.220972 | 0.356955 | 0.220972 | | |
| | 0.220972 | 0.220972 | 0.32811 | | |
| $(Os_j − Es_j)$, scaled | −2.11321 | 5.235849 | −2.81132 | −0.31132 | 0 |

$H_{FL} = (Os_j − Es_j)^T V^{−1}(Os_j − Es_j)$, the test statistic calculated for this single data set, is then estimated as 5.105 on 3 df, with the corresponding probability 0.1642. Again, this is identical to the equivalent Kruskal–Wallis result (see Supplementary materials, Example analyses file).

We then apply the same technique to each of the other time periods, add up the variance/covariance matrices and the scaled differences $(Os_j − Es_j)$ over all the time periods, and calculate our test statistic for incidental lesions using the same formula for H:

**All incidental lesions combined**

| | Group | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |
| Variance/ covariance matrix | 7.94280 | −2.73136 | −3.08816 | −2.12328 |
| | −2.73136 | 7.84534 | −2.78371 | −2.33028 |
| | −3.08816 | −2.78371 | 7.87077 | −1.99890 |
| | −2.12328 | −2.33028 | −1.99890 | 6.45246 |
| Inverse of matrix (groups 1–3) | 0.21738 | 0.12115 | 0.12814 | |
| | 0.12115 | 0.21327 | 0.12296 | |
| | 0.12814 | 1.12296 | 0.22082 | |
| $(Os_j − Es_j)$, scaled | −4.80715 | 3.15403 | −3.12041 | |

The test statistic for all incidental lesions combined is estimated as 7.045 on 3 df, with the corresponding probability 0.0705.

Finally, we sum up the variances, covariances, and scaled differences $(Os_j − Es_j)$ from the fatal and incidental analyses and use the same formula for H to calculate an overall probability for the difference in frequency of the lesion over the different treatment groups:

**All fatal and incidental lesions combined**

| | Group | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |
| Variance/ covariance matrix | 10.41310 | −3.53521 | −3.96680 | −2.91109 |
| | −3.53521 | 10.22765 | −3.61427 | −3.07817 |
| | −3.96680 | −3.61427 | 10.39491 | −2.81384 |
| | −2.91109 | −3.07817 | −2.81384 | 8.80309 |
| Inverse of matrix (groups 1–3) | 0.16111 | 0.08826 | 0.09217 | |
| | 0.08826 | 0.15819 | 0.08925 | |
| | 0.09217 | 0.08925 | 0.16240 | |
| $(Os_j − Es_j)$, scaled | −6.13052 | 2.00757 | −3.55635 | |

$H_{FLH}$ is estimated as 9.325 on 3 df, with the corresponding probability 0.0253.

Thus, using this analysis, we find a significant difference between the groups, with $P = 0.025$. This contrasts with the Peto analysis using presence of the lesion, which was only just significant ($P = 0.049$), the simple Kruskal–Wallis analysis, which was not significant ($P = 0.120$), and the Kruskal–Wallis analysis, stratified by week of death (but not by fatality), which was not significant ($P = 0.107$) either (see Tables 4 and 5).

## Using the Fry–Lee–Hamling test for dose-related trend

Inference in this kind of experiment is usually mainly determined by the significance of the dose-related trend. Let us work through the calculation of the trend probabilities for this set of data, using trend coefficients proportional to the dose levels ($d_j$) of 0, 10, 25, and 50 for groups 1–4, respectively.

To calculate the trend statistic, Z, from the above, we use the formula:

$$Z = \sum_s Ts / \sqrt{\left( \sum_s UsWs \right)}$$

where, with appropriate scaling,

$$Ts = \sum_{j=1}^{k} d_j(Os_j − Es_j),$$

$$Us = 4Vs/Ns^2,$$

$$Ws = \sum_{j=1}^{k} d_j^2 Ns_j − \left( \sum_{j=1}^{k} d_j Ns_j \right)^2 \Big/ Ns,$$

and:

$$Vs = \left( \sum_{j=1}^{k} \sum_{i=1}^{Ns_j} R_{ij}^2 − Ns\, E_.^2 \right) \Big/ (Ns − 1)$$

**Table 5** Kruskal–Wallis test

| | Group | | | | Total | Trend |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | | |
| *Yates's correction on* | | | | | | |
| **Dose (mg/kg/day)** | **0** | **10** | **25** | **50** | | |
| Kidney: chronic progressive nephropathy | | | | | | |
| N | 50 | 50 | 50 | 50 | 200 | |
| MRank | 89.84 | 102.49 | 96.09 | 113.58 | 100.50 | |
| Median | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | |
| $\chi^2$ | | 1.85 | 0.43 | 4.75 | 5.83 | 4.23 |
| P | | 0.1741 | 0.5125 | 0.0293[+] | 0.1204 | 0.0396[+] |
| Ex P | | 0.1807 | 0.5168 | 0.0292[+] | | |
| *Stratified by grouped week of death – Yates's correction on* | | | | | | |
| N | 50 | 50 | 50 | 50 | 200 | |
| ORank | 47.08 | 53.99 | 48.33 | 57.59 | 207.00 | |
| D | −4.69 | 2.20 | −3.07 | 5.56 | 0.00 | |
| $\chi^2$ | | 2.63 | 0.16 | 1.86 | 6.11 | 3.10 |
| P | | 0.1046 | 0.6852 | 0.1721 | 0.1066 | 0.0783[(+)] |
| *Yates's correction off* | | | | | | |
| N | 50 | 50 | 50 | 50 | 200 | |
| MRank | 89.84 | 102.49 | 96.09 | 113.58 | 100.50 | |
| Median | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | |
| $\chi^2$ | | 1.85 | 0.43 | 4.75 | 5.83 | 4.23 |
| P | | 0.1741 | 0.5125 | 0.0293[+] | 0.1204 | 0.0396[+] |
| Ex P | | 0.1807 | 0.5168 | 0.0292[+] | | |
| *Stratified by grouped week of death – Yates's correction off* | | | | | | |
| N | 50 | 50 | 50 | 50 | 200 | |
| ORank | 47.08 | 53.99 | 48.33 | 57.59 | 207.00 | |
| D | −4.69 | 2.20 | −3.07 | 5.56 | 0.00 | |
| $\chi^2$ | | 3.39 | 0.38 | 2.49 | 6.11 | 3.10 |
| P | | 0.0654[(+)] | 0.5391 | 0.1143 | 0.1066 | 0.0783[(+)] |

**Abbreviations:** Ex P, P calculated using exact methods; MRank, mean rank; ORank, observed rank; +, $P < 0.05$; (+), $P < 0.1$.

We have already calculated the values of $Os_j$, $Es_j$, and the sum of $R_{ij}^2$ when we were calculating the heterogeneity probability, and thus for the fatal analysis for week 56, we have:

Fatal lesions in week 56

| | Group | | | | Total |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | |
| $N_j$ | 48 | 46 | 49 | 48 | 191 |
| Sum of squares of ranks | 437772 | 419531.5 | 446892.25 | 465132.75 | 1769328.5 |
| $(Os_j − Es_j)$, scaled | −0.25131 | −0.24084 | −0.25654 | 0.74869 | |
| Dose = $d_j$ | 0 | 10 | 25 | 50 | |
| $d_j \times N_j$ | 0 | 460 | 1225 | 2400 | 4085 |
| $d_j^2 \times N_j$ | 0 | 4600 | 30625 | 120000 | 155225 |
| $Ts = d_j \times (Os_j − Es_j)$ | 0 | −2.40838 | −6.41361 | 37.43455 | 28.61257 |
| Ws | | | | | 67857.33 |
| Vs | | | | | 47.75 |
| Us | | | | | 0.005236 |

$Z_{FL} = Ts/\sqrt{(UsWs)}$, the test statistic calculated for this single time period, is then estimated as 1.5180, with the corresponding probability 0.1290.

As before, we then apply the same technique to each of the other time periods. By adding up the Ts and the UsWs values over all the time periods, we can calculate the test statistic for all fatal lesions as 1.8341 ($P = 0.0666$).

Using exactly the same approach for the incidental analysis and combining over the grouped time periods, the test statistic is 1.6783 ($P = 0.0933$). Finally, adding the Ts and UsWs values for the incidental analysis and the fatal analysis gives a value of $Z_{FLH} = 2.3749$ ($P = 0.0176$).

Thus, using this analysis, we find a significant trend across the groups, with $P = 0.018$. This contrasts with the Peto analysis using presence of the lesion at any severity level, which was not significant (exact $P = 0.073$, asymptotic $P = 0.069$; see the Peto analyses section, above), the simple Kruskal–Wallis analysis, which was marginally significant ($P = 0.040$) and the Kruskal–Wallis analysis, stratified by

**Table 6** Fry–Lee–Hamling test

| | Group | | | | Total | Trend |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | | |
| *Kruskal–Wallis Peto – Yates's correction on* | | | | | | |
| **Dose (mg/kg/day)** | **0** | **10** | **25** | **50** | | |
| **Kidney: chronic progressive nephropathy** | | | | | | |
| Fatal | | | | | | |
| N | 50 | 50 | 50 | 50 | 200 | |
| MRank | 1001.51 | 949.52 | 1037.49 | 941.48 | 982.50 | |
| Median | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | |
| ORank | 500.76 | 474.76 | 518.74 | 470.74 | 1965.00 | |
| D | −1.32 | −1.15 | −0.44 | 2.91 | 0.00 | |
| $\chi^2$ | | 0.19 | 0.00 | 1.39 | 3.74 | 3.36 |
| P | | 0.6619 | 0.9471 | 0.2392 | 0.2910 | 0.0666[(+)] |
| Ex P | | NS | NS | NS | | |
| Incidental | | | | | | |
| N | 48 | 48 | 47 | 44 | 187 | |
| MRank | 83.04 | 98.72 | 87.71 | 105.41 | 93.50 | |
| Median | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | |
| ORank | 42.63 | 50.68 | 44.09 | 49.61 | 187.00 | |
| D | −4.81 | 3.15 | −3.12 | 4.77 | 0.00 | |
| $\chi^2$ | | 3.97 | 0.06 | 1.93 | 7.05 | 2.82 |
| P | | 0.0464[+] | 0.8036 | 0.1643 | 0.0705[(*)] | 0.0933[(+)] |
| Ex P | | 0.0452[+] | 0.8977 | 0.8050 | | |
| Total | | | | | | |
| N | 50 | 50 | 50 | 50 | 200 | |
| MRank | 1086.77 | 1050.88 | 1125.66 | 1040.69 | 1076.00 | |
| Median | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | |
| $\chi^2$ | | 3.33 | 0.16 | 4.01 | 9.33 | 5.64 |
| P | | 0.0680[(+)] | 0.6927 | 0.0451[+] | 0.0253* | 0.0176[+] |
| *Kruskal–Wallis Peto – Yates's correction off* | | | | | | |
| Fatal | | | | | | |
| N | 50 | 50 | 50 | 50 | 200 | |
| MRank | 1001.51 | 949.52 | 1037.49 | 941.48 | 982.50 | |
| Median | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | |
| ORank | 500.76 | 474.76 | 518.74 | 470.74 | 1965.00 | |
| D | −1.32 | −1.15 | −0.44 | 2.91 | 0.00 | |
| $\chi^2$ | | 0.00 | 0.15 | 2.35 | 3.74 | 3.36 |
| P | | 0.9496 | 0.7031 | 0.1255 | 0.2910 | 0.0666[(+)] |
| Ex P | | NS | NS | NS | | |
| Incidental | | | | | | |
| N | 48 | 48 | 47 | 44 | 187 | |
| MRank | 83.04 | 98.72 | 87.71 | 105.41 | 93.50 | |
| Median | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | |
| ORank | 42.63 | 50.68 | 44.09 | 49.61 | 187.00 | |
| D | −4.81 | 3.15 | −3.12 | 4.77 | 0.00 | |
| $\chi^2$ | | 4.98 | 0.23 | 2.64 | 7.05 | 2.82 |
| P | | 0.0257[+] | 0.6324 | 0.1044 | 0.0705[(*)] | 0.0933[(+)] |
| Ex P | | 0.0452[+] | 0.8977 | 0.8050 | | |
| Total | | | | | | |
| N | 50 | 50 | 50 | 50 | 200 | |
| MRank | 1086.77 | 1050.88 | 1125.66 | 1040.69 | 1076.00 | |
| Median | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | |
| $\chi^2$ | | 4.16 | 0.36 | 4.83 | 9.33 | 5.64 |
| P | | 0.0413[+] | 0.5488 | 0.0279[+] | 0.0253* | 0.0176[+] |

**Abbreviations:** MRank, mean rank; NS, not significant; ORank, observed rank; + and *, $P < 0.05$; (+) and (*), $P < 0.1$.

**Table 7** Poly-$k$ test

| | Group | | | | Total | Trend |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | | |
| **Poly-3 analyses – Yates's correction on** | | | | | | |
| Dose (mg/kg/day) | 0 | 10 | 25 | 50 | | |
| **Kidney: chronic progressive nephropathy** | | | | | | |
| ~severe | | | | | | |
| n | 1 | 2 | 1 | 9 | 13 | |
| E | 3.32 | 3.13 | 3.41 | 3.13 | 13.00 | |
| E(2) | | 1.46 | 1.01 | 4.85 | | |
| $\chi^2$ | | 0.00 | 0.49 | 6.08 | 16.00 | 11.79 |
| P | | 0.9588 | 0.4855 | 0.0137+ | 0.0011** | 0.0006+++ |
| ~severe or ~marked | | | | | | |
| n | 1 | 2 | 3 | 15 | 21 | |
| E | 5.23 | 4.94 | 5.52 | 5.31 | 21.00 | |
| E(2) | | 1.46 | 2.05 | 8.06 | | |
| $\chi^2$ | | 0.00 | 0.21 | 12.84 | 27.48 | 23.62 |
| P | | 0.9588 | 0.6464 | 0.0003+++ | 0.0000*** | 0.0000+++ |
| At least ~moderate | | | | | | |
| n | 3 | 3 | 6 | 17 | 29 | |
| E | 7.25 | 6.80 | 7.60 | 7.35 | 29.00 | |
| E(2) | | 2.90 | 4.60 | 10.07 | | |
| $\chi^2$ | | 0.12 | 0.40 | 10.87 | 21.34 | 19.21 |
| P | | 0.7317 | 0.5279 | 0.0010+++ | 0.0001*** | 0.0000+++ |
| At least ~slight | | | | | | |
| n | 5 | 7 | 9 | 19 | 40 | |
| E | 9.95 | 9.44 | 10.52 | 10.09 | 40.00 | |
| E(2) | | 5.84 | 7.19 | 12.08 | | |
| $\chi^2$ | | 0.17 | 0.58 | 9.62 | 14.70 | 13.62 |
| P | | 0.6799 | 0.4446 | 0.0019++ | 0.0021** | 0.0002+++ |
| Any grade | | | | | | |
| n | 17 | 24 | 19 | 23 | 83 | |
| E | 20.59 | 19.58 | 21.80 | 21.04 | 83.00 | |
| E(2) | | 19.98 | 18.51 | 20.22 | | |
| $\chi^2$ | | 2.44 | 0.00 | 1.00 | 4.29 | 0.35 |
| P | | 0.1181 | 0.9955 | 0.3184 | 0.2313 | 0.5526 |
| **Poly-3 analyses – Yates's correction off** | | | | | | |
| ~severe | | | | | | |
| n | 1 | 2 | 1 | 9 | 13 | |
| E | 3.32 | 3.13 | 3.41 | 3.13 | 13.00 | |
| E(2) | | 1.46 | 1.01 | 4.85 | | |
| $\chi^2$ | | 0.41 | 0.00 | 7.86 | 16.00 | 11.79 |
| P | | 0.5220 | 0.9854 | 0.0051++ | 0.0011** | 0.0006+++ |
| ~severe or ~marked | | | | | | |
| n | 1 | 2 | 3 | 15 | 21 | |
| E | 5.23 | 4.94 | 5.52 | 5.31 | 21.00 | |
| E(2) | | 1.46 | 2.05 | 8.06 | | |
| $\chi^2$ | | 0.41 | 0.94 | 14.92 | 27.48 | 23.62 |
| P | | 0.5220 | 0.3315 | 0.0001+++ | 0.0000*** | 0.0000+++ |
| At least ~moderate | | | | | | |
| n | 3 | 3 | 6 | 17 | 29 | |
| E | 7.25 | 6.80 | 7.60 | 7.35 | 29.00 | |
| E(2) | | 2.90 | 4.60 | 10.07 | | |
| $\chi^2$ | | 0.01 | 0.97 | 12.62 | 21.34 | 19.21 |
| P | | 0.9349 | 0.3252 | 0.0004+++ | 0.0001*** | 0.0000+++ |

(*Continued*)

**Table 7** (*Continued*)

| | Group | | | | Total | Trend |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | | |
| At least ~slight | | | | | | |
| n | 5 | 7 | 9 | 19 | 40 | |
| E | 9.95 | 9.44 | 10.52 | 10.09 | 40.00 | |
| E(2) | | 5.84 | 7.19 | 12.08 | | |
| $\chi^2$ | | 0.53 | 1.12 | 11.18 | 14.70 | 13.62 |
| P | | 0.4681 | 0.2906 | 0.0008+++ | 0.0021** | 0.0002+++ |
| ***Poly-3 analyses – Yates's correction off*** | | | | | | |
| Any grade | | | | | | |
| n | 17 | 24 | 19 | 23 | 83 | |
| E | 20.59 | 19.58 | 21.80 | 21.04 | 83.00 | |
| E(2) | | 19.98 | 18.51 | 20.22 | | |
| $\chi^2$ | | 3.19 | 0.05 | 1.48 | 4.29 | 0.35 |
| P | | 0.0742(+) | 0.8310 | 0.2239 | 0.2313 | 0.5526 |

**Notes:** Decedents after start of terminal kill at week 105 treated as kills.
**Abbreviations:** E, expected value considering all groups; E(2), expected value considering only a single treated group and the control group; +++ and ***, $P < 0.001$; ++ and **, $P < 0.01$; + and *, $P < 0.05$; (+) and (*), $P < 0.1$.

**Table 8** Modified poly-*k* test

| | Group | | | | Total | Trend |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | | |
| ***Modified poly-3 analyses – Yates's correction on*** | | | | | | |
| **Dose (mg/kg/day)** | **0** | **10** | **25** | **50** | | |
| **Kidney: chronic progressive nephropathy** | | | | | | |
| ~severe | | | | | | |
| n | 0.83 | 1.56 | 0.85 | 7.01 | 13 | |
| E | 2.67 | 2.38 | 2.81 | 2.38 | | |
| E(2) | | 1.12 | 0.86 | 3.69 | | |
| $\chi^2$ | | 0.01 | 0.69 | 6.11 | 16.82 | 12.17 |
| P | | 0.9253 | 0.4068 | 0.0134+ | 0.0008*** | 0.0005+++ |
| ~severe or ~marked | | | | | | |
| n | 0.83 | 1.56 | 2.61 | 12.57 | 21 | |
| E | 4.36 | 3.88 | 4.84 | 4.49 | | |
| E(2) | | 1.12 | 1.81 | 6.80 | | |
| $\chi^2$ | | 0.01 | 0.13 | 14.65 | 31.85 | 27.26 |
| P | | 0.9253 | 0.7154 | 0.0001+++ | 0.0000*** | 0.0000+++ |
| At least ~moderate | | | | | | |
| n | 2.49 | 2.34 | 5.22 | 14.32 | 29 | |
| E | 6.08 | 5.35 | 6.68 | 6.26 | | |
| E(2) | | 2.26 | 4.04 | 8.52 | | |
| $\chi^2$ | | 0.20 | 0.32 | 12.14 | 23.92 | 21.49 |
| P | | 0.6575 | 0.5703 | 0.0005+++ | 0.0000*** | 0.0000+++ |
| At least ~slight | | | | | | |
| n | 4.15 | 5.51 | 7.90 | 16.01 | 40 | |
| E | 8.30 | 7.47 | 9.27 | 8.54 | | |
| E(2) | | 4.58 | 6.36 | 10.22 | | |
| $\chi^2$ | | 0.11 | 0.51 | 10.64 | 15.99 | 14.77 |
| P | | 0.7354 | 0.4750 | 0.0011++ | 0.0011** | 0.0001+++ |
| Any grade | | | | | | |
| n | 14.12 | 18.96 | 16.71 | 19.53 | 83 | |
| E | 17.03 | 15.40 | 19.09 | 17.79 | | |
| E(2) | | 15.71 | 16.29 | 17.19 | | |
| $\chi^2$ | | 2.51 | 0.00 | 0.97 | 4.53 | 0.41 |
| P | | 0.1130 | 0.9652 | 0.3244 | 0.2099 | 0.5231 |

(*Continued*)

**Table 8** (*Continued*)

| | Group | | | | Total | Trend |
|---|---|---|---|---|---|---|
| | **1** | **2** | **3** | **4** | | |
| *Modified poly-3 analyses – Yates's correction off* | | | | | | |
| ~severe | | | | | | |
| n | 0.83 | 1.56 | 0.85 | 7.01 | 13 | |
| E | 2.67 | 2.38 | 2.81 | 2.38 | | |
| E(2) | | 1.12 | 0.86 | 3.69 | | |
| $\chi^2$ | | 0.40 | 0.00 | 8.47 | 16.82 | 12.17 |
| P | | 0.5250 | 0.9855 | 0.0036++ | 0.0008*** | 0.0005+++ |
| ~severe or ~marked | | | | | | |
| n | 0.83 | 1.56 | 2.61 | 12.57 | 21 | |
| E | 4.36 | 3.88 | 4.84 | 4.49 | | |
| E(2) | | 1.12 | 1.81 | 6.80 | | |
| $\chi^2$ | | 0.40 | 0.93 | 17.55 | 31.85 | 27.26 |
| P | | 0.5250 | 0.3336 | 0.0000+++ | 0.0000*** | 0.0000+++ |
| At least ~moderate | | | | | | |
| n | 2.49 | 2.34 | 5.22 | 14.32 | 29 | |
| E | 6.08 | 5.35 | 6.68 | 6.26 | | |
| E(2) | | 2.26 | 4.04 | 8.52 | | |
| $\chi^2$ | | 0.01 | 0.96 | 14.54 | 23.92 | 21.49 |
| P | | 0.9354 | 0.3263 | 0.0001+++ | 0.0000*** | 0.0000+++ |
| At least ~slight | | | | | | |
| n | 4.15 | 5.51 | 7.90 | 16.01 | 40 | |
| E | 8.30 | 7.47 | 9.27 | 8.54 | | |
| E(2) | | 4.58 | 6.36 | 10.22 | | |
| $\chi^2$ | | 0.53 | 1.12 | 12.75 | 15.99 | 14.77 |
| P | | 0.4687 | 0.2904 | 0.0004+++ | 0.0011** | 0.0001+++ |
| *Modified poly-3 analyses – Yates's correction off* | | | | | | |
| Any grade | | | | | | |
| n | 14.12 | 18.96 | 16.71 | 19.53 | 83 | |
| E | 17.03 | 15.40 | 19.09 | 17.79 | | |
| E(2) | | 15.71 | 16.29 | 17.19 | | |
| $\chi^2$ | | 3.51 | 0.05 | 1.57 | 4.53 | 0.41 |
| P | | 0.0611(+) | 0.8293 | 0.2100 | 0.2099 | 0.5231 |

**Notes:** Decedents after start of terminal kill at week 105 treated as kills.
**Abbreviations:** E, expected value considering all groups; E(2), expected value considering only a single treated group and the control group; +++ and ***, $P < 0.001$; ++ and **, $P < 0.01$; + and *, $P < 0.05$; (+) and (*), $P < 0.1$.

week of death (but not by fatality), which was not significant ($P = 0.078$) (see Table 5).

## Conclusion

The analysis of graded lesions in carcinogenicity studies has not been properly addressed. As for the Peto test, we need to take into account the context of the tumor ("fatal" or "incidental") and the time of death of the animal. However, we also need to take into account the severity of the lesion observed. The extension to the Fry–Lee test put forward here allows for all these relevant factors. Note that the "rolling cuts" method, in which repeated Peto tests are conducted for different severity-level cutoffs is also a good methodology to use, but has problems related to multiple testing.

Our new test can also be used where tumors form a severity gradation. If, for a specific organ and tissue type, it is considered that an adenocarcinoma develops from an adenoma, we can analyze these tumors in a single analysis, setting scores of 0 = no tumor, 1 = adenoma, and 2 = adenocarcinoma. Similarly, where focal hyperplasia is considered a precursor to the tumor, and is graded on a 5-point scale (1 = minimal to 5 = severe), we can form a single variable by adding grades 6 = adenoma and level 7 = adenocarcinoma, and consider focal hyperplasias and tumors together in a single analysis.

The Fry–Lee–Hamling test can be regarded as a more general test than the others. For lesions recorded only as present or absent, the test for heterogeneity provides identical results to those from the Peto test when context of observation

and time of death are taken into account, to those from the stratified chi-squared test when all lesions are incidental, and to those from the simple chi-squared test when age adjustment is not attempted. Also, for lesions recorded by severity, the test provides identical results to the stratified Kruskal–Wallis (Fry–Lee) test when all lesions are incidental and to those from the Kruskal–Wallis test when age adjustment is not attempted. The test for trend is similarly equivalent.

Note that where it has proved to be impossible for the pathologist to determine the context of observations, another form of test – the poly-$k$ test[9,10] – has been suggested. Our reservations about this test have already been published,[11] and it is worth noting that the US Food and Drug Administration requires pathologists to produce a table giving factors contributory to death. This suggests that there is no inherent problem in providing a context of observation for lesions seen in carcinogenicity studies. It is also interesting to note that analyses of the data by the poly-$k$ method, with $k = 3$, produced nonsignificant answers ($P = 0.2313$ for the standard analysis, see Table 7; $P = 0.2099$ for the modified form) whereas our analysis gave a significant answer ($P = 0.0253$).

## Disclosure

## References

1. Peto R, Pike MC, Day NE, et al. Guidelines for simple, sensitive significance tests for carcinogenic effects in long-term animal experiments. *IARC Monogr Eval Carcinog Risk Chem Hum Suppl*. 1980;Suppl 2:311–426.
2. Gart JJ, Krewski D, Lee PN, Tarone RE, Wahrendorf J. Statistical methods in cancer research. Volume III – The design and analysis of long-term animal experiments. *IARC Sci Publ*. 1986;(79):1–219.
3. Kruskal WH, Wallis WA. Use of ranks in one-criterion variance analysis. *J Am Stat Assoc*. 1952;47(260):583–621.
4. Marascuilo LA, McSweeney M. Nonparametric post hoc comparisons for trend. *Psychol Bull*. 1967;67(6):401–412.
5. Cuzick J. The efficiency of the proportions test and the logrank test for censored survival data. *Biometrics*. 1982;38(4):1033–1039.
6. Fry JS, Lee PN. Stratified rank tests [letter]. *Appl Stat*. 1988;37: 264–266.
7. Heimann G, Neuhaus G. Permutational distribution of the log-rank statistic under random censorship with applications to carcinogenicity assays. *Biometrics*. 1998;54(1):168–184.
8. Fry JS, Lee PN. A stratified Wilcoxon-type test for trend. *Stat Med*. 1991;10(5):799–800.
9. Bailer AJ, Portier CJ. Effects of treatment-induced mortality and tumor-induced mortality on tests for carcinogenicity in small samples. *Biometrics*. 1988;44(2):417–431.
10. Bieler GS, Williams RL. Ratio estimates, the delta method, and quantal response tests for increased carcinogenicity. *Biometrics*. 1993;49(3):793–801.
11. [No authors listed]. Statistical methods for carcinogenicity studies. *Toxicol Pathol*. 2002;30(3):403–414.

## Supplementary materials

## Calculation of grouped periods of death for the incidental part of the Peto test

The algorithm used was developed as a routine method to produce reasonable intervals for many standard types of experiment. It is defined as follows:

Define a score corresponding to each experiment week:

1. for studies up to 104 weeks

| Weeks | Score |
|-------|-------|
| 1 to 52 | 1 |
| 53 to 78 | 2 |
| 79 to 91 | 4 |
| 92 to 104 | 8 |

2. for studies over 104 weeks (of length T)

| Weeks | Score |
|-------|-------|
| 1 to (T−52) | 1 |
| (T−51) to (T−26) | 2 |
| (T−25) to (T−13) | 4 |
| (T12) to T | 8 |

Create the total score, S, by summing scores corresponding to each study week.

Compute the total number of time groups, N, by dividing S by 52 and rounding up, if necessary, to the next-higher whole number.

If $S \leq 156$, create N consecutive time groups, each of score "close" to S/N. If $S > 156$, make the first time group 1−52 and create $N - 1$ further consecutive time groups each of score "close" to $(S - 52)/(N - 1)$.

The above applies to experiments in which all animals die, or when there is a single killing at the end of the study, in which case an additional level of the time-group variable is allocated for the killed animals. If there are interim killings, a further level of the time-group variable is created for each, and time groups for decedents are created by dividing periods between killings using the same rules as above (except that if the total score between killings is >156, no first grouping of weeks 1−52 is forced). Note that one interim killing may run onto several consecutive weeks (eg, weeks 53−55). If so, the last death period before the killing will end at the highest of these weeks (here 55), with the next death period starting at the next week. This allows for deaths occurring during the killing period.

## The exact trend test

An exact test for dose-related trend can be calculated using an extension to the Fisher exact test. The test is based on calculating the exact one-tailed probability of achieving a score, S, as or more extreme than actually observed, given fixed marginal totals, where:

$$S = \sum_{j=1}^{k} a_j d_j$$

$a_j$ = number of animals at risk in group j with the condition present
$d_j$ = dose level applied to group j
k = number of dose groups (including any control group)

Consider the data in table form, as follows:

| | Dose group | | | | | Total |
|---|---|---|---|---|---|---|
| | I | 2 | . . . | k − I | k | |
| At risk, with the condition | $a_1$ | $a_2$ | . . . | $a_{k-1}$ | $a_k$ | $n_1$ |
| At risk, without the condition | $c_1$ | $c_2$ | . . . | $c_{k-1}$ | $c_k$ | $n_0$ |
| Total | $m_1$ | $m_2$ | . . . | $m_{k-1}$ | $m_k$ | N |

For the exact trend test, the marginal totals are considered fixed. Thus, for defined marginal totals, choosing values for $a_1$ to $a_{k-1}$ determines the other entries in this table.

The probability of a given table comes from the multinomial distribution:

$$p = (m_1! \; m_2! \; ... \; m_k! \; n_0! \; n_1!)/(a_1! \; a_2! \; ... \; a_k! \; c_1! \; c_2! \; ... \; c_k! \; N!)$$

where:

$c_j$ = number of animals at risk in group j with the condition absent

$m_j$ = number of animals at risk in group j

$n_1$ = total number of animals at risk with the condition present

$n_0$ = total number of animals at risk with the condition absent

$N$ = total number of animals at risk

The problem then resolves into enumerating all the possible tables that have the same marginal totals and calculating their scores and probabilities.

The counting algorithm we use is as follows. First, imagine the problem in terms of a series of posts, with each group having one post, and the height of the post the marginal total associated with that group. The posts are ordered by the dose values associated with them. Then we take a collection of rings equal to the number of animals with the condition. Placing the rings over the posts is equivalent to setting up one of the possible tables in our enumeration; we just have to ensure that all the rings are placed and that the number of rings on each post never exceeds the "height" for that post (the total number of animals originally in that group).

We first define our starting position, which is the table with the smallest possible score, by placing the maximum number of rings on the first post, then filling up the second post and so on until all the rings are used up. The score for this table is calculated by multiplying the numbers of rings on each post by the associated dose value and adding up over all the posts. The probability for this table comes from the multinomial distribution, as given above.

The algorithm can be stated briefly as:

### Algorithm

We are trying to find the significance of the observed table by calculating the sum of the probabilities for tables, such that their score is less than or equal to the score for the observed table. If the sum of probabilities is greater than the lowest significance level in which we are interested (minP), we then stop the process described below.

1  Find the minimal legal starting position, filling the posts from left to right.
   Calculate the probability (P) and score for this position.
   Add P into the total probability (TP).
   IF TP is greater than minP, STOP.
   Identify the starting post as the highest post with rings present – IPost.
   IPost will change during the course of the algorithm.
2  IF IPost < 1, STOP.
   IF IPost < highest post and there are rings on IPost, THEN
Look for the shortest legal moves of a ring from IPost to a higher post where the score is less than or equal to the observed score.
      If possible
         Do the move, and store the new position together with the associated score and probability against each post from the original IPost to the post the ring has moved to.
      Identify the post the ring has moved to as the post of interest (IPost).
         Go to 2.
      If not possible
         Go to 3.
3  Identify the next-lower post as the post of interest (IPost).
         **Go to 2.**

Thus, we look at the post with the highest dose level that has rings on it and see if we can do a legal move, ie, to move one of the rings to the next-highest post possible. This move will then produce a new legal table with a higher score than we started with. If no legal move is possible, we try to move a ring from the next-lower post. When there are no posts left, we stop.

Where we have found a new legal table, we can calculate its score and its probability based on the move itself. That is:

new score = old score − dose for old post + dose for new post

and where:

pold = rings on old post, now reduced by one

pnew = rings on new post, now increased by one

aold = vacant positions for rings on old post, now increased by one

anew = vacant positions for rings on new post, now decreased by one

Then we can calculate the probability for the new table (xpnew) from these numbers and the probability for the old table (xpold) by:

xpnew = xpold ((anew + 1)/aold))/(pnew/(pold + 1))

Where the score for the new table is equal to the observed score (S), we add the probability into a running probability for tables with scores equal to the observed score.

Where the score is strictly less than S, we add the probability into a running probability for tables with scores less than the observed score.

Most parts of the algorithm are now in place, but when we have moved a ring to the highest post, we must be able to backtrack to a previous position. Fortunately, all we have to do is to remember the position of the rings, the score, and the probability against each post. When we move a ring, we update the position, score, and probability for the old post, the new post, and the posts in between, but not for posts with lower values.

When we cannot move the ring to a higher position, we return to the position remembered for the next-lower post and see if a legal move can be made.

This process should end up with us having a probability that the observed score is achieved (obsp) and the probability of achieving a score of strictly less than the observed score (probT). From these, we can calculate one-sided probabilities for the trend, or a two-sided probability which we take as:

2 × minimum (probT + obsp, (1 − probT) + obsp)

A version of this algorithm, written in Visual Basic for Applications (VBA), is included in the Excel code provided in the Supplementary file, Example analyses.

## Example

Let us have four groups, with three animals per group. Let the trend be 0, 1, 5 and 25 for the 4 groups respectively. Let us take our observed number of animals with the condition as (2,2,1,0) in groups 1, 2, 3, and 4 respectively. This will have an observed score of $S = (2 \times 0) + (2 \times 1) + (1 \times 5) + (0 \times 25) = 7$. In this example, let us look for a negative trend, ie, for tables that have scores less than or equal to 7. Let us also set the minimum probability to 1, ie, we will look at all tables that have the relevant scores.

As we have a maximum of three animals per group, our minimum starting position is (3,2,0,0) with score = 2 and probability = (3! 3! 3! 3! 5! 7!)/(3! 2! 3! 3! 0! 1! 0! 0! 12!) = 0.003788.

IPost starts at 2.
Try to move an animal from group 2 to group 3.
New position = (3,1,1,0)
New score = old score − dose for old post + dose for new post = 2 + 5 − 1 = 6
This score is less than the observed score of 7, so we can do the move

xpnew = xpold ((anew + 1)/aold))/(pnew/(pold + 1))

    = 0.003788 × ((2 + 1)/2)/(1/(1 + 1)) = 0.011364

Add probability into the total: xptot = 0.003788 + 0.011364 = 0.015152

Remember this position for posts 2 and 3; post 1still has starting position (3,2,0,0).


IPost becomes 3.

Try to move an animal from group 3 to group 4.

New position = (3,1,0,1)

    New score = old score − dose for old post + dose for new post = 6 + 25 − 5 = 26

    This score is greater than the observed score of 7, so do not do move.


IPost goes down from 3 to 2 with remembered position (3,1,1,0).

Try to move an animal from post 2 to post 3.

New position = (3,0,2,0)

    New score = old score − dose for old post + dose for new post = 6 + 5 − 1 = 10

    This score is greater than the observed score of 7, so do not do move.


IPost goes down from 2 to 1 with remembered position (3,2,0,0).

Try to move an animal from post 1 to post 2.

New position = (2,3,0,0)

    New score = old score − dose for old post + dose for new post

        = 2 + 1 − 0 = 3, which is <7;

    xpnew = xpold ((anew + 1)/aold))/(pnew/(pold + 1)) = 0.003788 × ((1)/1)/(3/(3)) = 0.003788

    Add probability into the total: xptot = 0.015152 + 0.003788 = 0.018939

Record position against posts 1 and 2.

IPost becomes 2 with position (2,3,0,0).


Try to move an animal from post 2 to post 3.

New position = (2,2,1,0)

    New score = old score − dose for old post + dose for new post = 3 + 5 − 1 = 7

    This score is equal to the observed score of 7, so do move

    xpnew = xpold ((anew + 1)/aold))/(pnew/(pold + 1)) = 0.003788 × ((3)/1)/(1/(3)) = 0.0341

    Add probability into the total: xptot = 0.018939 + 0.0340909 = 0.053030

Record position against posts 2 and 3 − post 1 still has position (2,3,0,0).

IPost becomes 3 with position (2,2,1,0).


Try to move an animal from post 3 to post 4.

New position = (2,2,0,1)

    New score = old score − dose for old post + dose for new post = 7 + 25 − 5 = 27

    This score is greater than the observed score of 7, so do not do move

IPost goes down from 3 to 2 with position (2,2,1,0).


Try to move an animal from post 2 to post 3.

New position = (2,1,2,0)

    New score = old score − dose for old post + dose for new post = 7 + 5 − 1 = 11

    This score is greater than the observed score of 7, so do not do move

IPost goes down from 2 to 1 with position (2,3,0,0).

Try to move an animal from post 1 to post 2.
New position = (1,4,0,0)
This would give too many animals on post 2, so we try again, moving to a higher post.

Try to move an animal from post 1 to post 3.
New position = (1,3,1,0)
    New score = old score − dose for old post + dose for new post
        = 3 + 5 − 0 = 8 which is >7

IPost goes down from 1 to 0 and therefore we stop.

Thus, we found four legitimate tables with scores less than or equal to the observed table, with total probability = 0.053030.

Note that in the VBA code supplied in the Supplementary file, we deal with logs of the probabilities rather than the actual probabilities, in order to keep rounding errors under control.

## Equivalence of the test statistics used in calculating Kruskal–Wallis statistics

The formulation often suggested for calculation of Kruskal–Wallis statistics when there are tied observations is:

$$H = \left[ 12 \sum_{j=1}^{k} (R_j^2 / N_j) \, / \, N(N+1) - 3(N+1) \right] \Big/ C$$

where N is the total number of animals, k is the number of groups, $N_j$ is the number of animals in group j, $R_j$ is the sum of ranks of animals in group j, and

$$C = 1 - \left( \sum t^3 - \sum t \right) \Big/ (N^3 - N)$$

where t is the number of tied observations for a given tied rank. H is asymptotically distributed as chi-squared on k − 1 degrees of freedom (df).

This formula can be re-written without the use of tied ranks as:

$$H' = \sum_{j=1}^{k} (O_j - E.)^2 \, N_j \Big/ V$$

where $O_j$ is the mean rank for group j, E. is the expected rank assuming no difference between groups, and V is the sample variance assuming no difference between the groups. Defining $R_{ij}$ as the rank for animal i in group j, we have:

$$O_j = \sum_{i=1}^{N_j} R_{ij} \, / \, N_j$$

$$E_j = (N + 1) \, / \, 2 \ \text{ and}$$

$$V = (\sum_{j=1}^{k} \sum_{i=1}^{N_j} R_{ij}^2 - NE.^2) \, / \, (N-1)$$

H′ is also distributed asymptotically as chi-squared on k − 1 df.

Let us start with the H′ equation, replacing $O_j$ and E. with the values given above:

$$H' = \left[ \sum_{j=1}^{k} (R_j / N_j - (N+1)/2)^2 \, N_j \right] / V$$

$$= \left[ \sum_{j=1}^{k} \left( R_j^2 / N_j + (N+1)^2 \, N_j / 4 - (N+1)R_j \right) \right] / V$$

Now the sum of ranks or tied ranks is the same, $N(N + 1)/2$, and the sum of $N_j$ is N. This gives:

$$H' = \left[ \sum_{j=1}^{k} (R_j^2 / N_j) - N(N+1)^2/4 \right] / V$$

Thus, the problem reduces to showing that V is equivalent to $C(N(N + 1)/12)$.

Let us start with the situation of no ties. Then:

$$V = (\sum_{i=1}^{N} R_i^2 - NE.^2) / (N-1)$$
$$= \left( N(N+1)(2N+1)/6 - N(N+1)^2/4 \right)/(N-1)$$
$$= N(N+1)/12$$

Now let there be m different values for the variable of interest with $t_j$ tied values for each of the m values. Now we have:

$$V = (\sum_{j=1}^{m} t_j R_j^2 - NE.^2) / (N-1)$$

Consider the ith set of tied ranks, with $t_i \geq 1$. Then consider a second situation in which each of those $t_i$ observations has nontied values. We will examine the difference in the variance between these two situations.

In the first situation, (tied ranks), if there are S observations smaller than those in the ith set of ranks, then the rank assigned to each of the $t_i$ values is $S + (t_i + 1)/2$. The value of the summation term within the variance equation is then:

$$t_i \left( S + (t_i +1)/2 \right)^2$$

In the second situation (no tied ranks), the value would be:

$$\sum_{k=1}^{t_i} (S+k)^2 = t_i S^2 + 2S \sum_{k=1}^{t_i} k + \sum_{k=1}^{t_i} k^2$$

The difference between these two values (with tied ranks – without tied ranks) is then:

$$= t_i \left( S^2 + S(t_i +1) + (t_i +1)^2/4 \right) - t_i \left( S^2 + S(t_i + 1) + (t_i + 1)(2t_i + 1)/6 \right)$$
$$= t_i (t_i +1)^2/4 - t_i (t_i +1)(2t_i + 1)/6$$
$$= t_i (t_i +1) (3t_i +3 - 4t_i -2)/12$$
$$= -(t_i^3 - t_i)/12$$

Thus, we have:

$$\sum R_{ij}^2 = N(N+1)(2N+1)/6 - \sum_{j=1}^{m} (t_j^3 - t_j)/12$$

and hence:

$$V = \left( N(N+1)(2N+1)/6 - \sum_{j=1}^{m} (t_j^3 - t_j)/12 - N(N+1)^2/4 \right) \bigg/ (N-1)$$

$$= \left( N(N+1)) \; (N-1) - \sum_{j=1}^{m} (t_j^2 - t_j) \right) \bigg/ 12 \, (N\text{-}1)$$

$$= N(N+1) \; (N-1) \left( 1 - \sum_{j=1}^{m} (t_j^3 - t_j) \; / \; (N^3 - N) \right) \bigg/ 12 \; (N-1)$$

$$= N(N+1) \; C/12$$

as required.

## Example analyses

http://www.dovepress.com/cr_data/supplementary_file_43535.xlsm