



Published in final edited form as:

Hum Ecol Risk Assess. 2009 September 1; 15(5): 858–875. doi:10.1080/10807030903153360.

Applications of the Poly-K Statistical Test to Life-Time Cancer Bioassay Studies

Mulugeta Gebregziabher¹ and David Hoel²

¹Medical University of South Carolina, Department of Biostatistics, Bioinformatics and Epidemiology, Charleston, SC, USA.

²Medical University of South Carolina, Department of Biostatistics, Bioinformatics and Epidemiology, Charleston, SC, USA

Abstract

The statistical analysis of cancer bioassay data has historically depended on the pathological determination of the experimental animal's cause of death. The poly-k statistical test has provided a method of statistical analysis of animal bioassay data without the need for cause of death information. The test has been shown to have good statistical properties in the typical 2-year cancer bioassay. However, while the poly-k test has been applied to chronic lifetime animal studies, it has not been formally evaluated with respect to the operating characteristics of this statistical test when applied to such studies. Thus, our objective is to assess the performance of the poly-k test for lifetime studies and to make comparisons with other tests. We observed in one recent lifetime study of the gasoline additive MTBE that the application of the poly-k test was not statistically robust. Simulation studies were subsequently conducted for a limited number of scenarios of lifetime cancer bioassays. These simulations showed that the poly-k test is not statistically robust for testing effect of increasing dose in some lifetime cancer studies.

Keywords

logistic regression score test; logrank test; survival adjusted trend tests; type-I error; re-weighted poly-k test; power

INTRODUCTION

The statistical analysis of animal cancer studies has enjoyed a long and evolving history. One reason why the topic has been important is the fact that the toxicological testing issue is basic to the use of data for extrapolating from animal risk to humans. For instance, our data example concerns the analysis of the risk of the gasoline additive chemical MTBE on Sprague-Dawley rats, which has been a point of policy debate due to its implication of risk of cancer to humans. Methods using basic competing risk methodologies were generally used until it was established that for incidental tumors, tumors that are occult and believed to not be a cause of death, and differential survivals among experimental groups could lead to misleading statistical interpretations of the experimental data (Hoel and Walburg 1972). For tumors that were considered to be a cause of death in tumor bearing animals, a traditional survival analysis method such as the Cox proportional hazards analysis, was generally accepted along with a Kaplan-Meier estimator of the probability of mortality as a function of time with regard to the

Address correspondence to Mulugeta Gebregziabher, Medical University of South Carolina, Department of Biostatistics, Bioinformatics and Epidemiology, 135 Cannon St., Charleston, SC 29425, Tel: 843-876-1112, Fax: 843-876-1126, gebregz@musc.edu.

tumor of interest. For incidental tumors Hoel and Walburg provided an alternative to Kaplan-Meier estimation as well as an interval method for testing treatment-control differences (Hoel and Walburg 1972).

In testing for treatment effects of incidental tumors (tumors that can be observed only if the animal died due to causes other than the tumor of interest), an improved method was developed by Dinse and Lagakos (Dinse and Lagakos 1983), who applied a logistic regression analysis that included time as a covariate. Peto (Peto *et al.* 1980) further developed methods in which the pathologist determined for each specific animal how likely the tumor of interest was the cause of death in tumor-bearing animals and developed a method for the overall testing of treatment effects for a given tumor type where there was a mixture of lethal and incidental effects on the mortality of the animals. All of these methods required a determination by the pathologist as to whether each animal's cancer was a likely contributor to the animal's death or was primarily independent of the animal's death.

Bailer and Portier (Bailer and Portier 1988; Portier and Bailer 1989) next developed a method of weighting an animal's time at risk, thereby avoiding the need for the pathologist determination of lethality. Their test, the “poly-k” (Portier *et al.* 1986), was based on observations from an extensive analysis of a large database of tumors from control rodents in a typical 2-year bioassay. They found that general survival time for most tumor types could be adequately modeled by a two parameter Weibull distribution. An estimate of the Weibull shape parameter, k as in poly-k, was then used as the power “ k ” in the poly-k weighting scheme taking values 1 to 5 with 3 being typical. The same paper uses a lifetime study of Fischer 344 control rats in an asbestos exposure study to support the applicability of the Weibull distribution to lifetime studies.

The National Toxicology Program (NTP) has adopted the poly-k method of analysis for its cancer bioassay program and it has been discussed in some detail by a number of statisticians concerned with its adoption as U.S. Food and Drug Administration (FDA) policy (STP Peto Group 2002). One concern with the test has been the assumption of the particular power for the weighting of the time at risk for a given animal with the default being $k = 3$. It has been reported that the test is quite robust over various values of k , so the choice of k is not of great concern (Bailer and Portier 1988; Portier and Bailer 1989; Dinse 1994). In the NTP applications, the animals are sacrificed at 2 years. However, some researchers have also applied the test to lifetime studies, as was recently done, *e.g.*, in an analysis of the cancer effects of aspartame (Belpoggi *et al.* 1995). But tests proposed for a 2-year study may not be appropriate for a lifetime study and the naïve application of tests proposed for 2-year studies to lifetime studies is not appropriate. So, our goal is to make an extensive assessment of the performance of the poly-k test in lifetime studies and to make comparisons with other tests based on whether or not the tumor is the cause of death.

AN EXAMPLE

A study of the gasoline additive MTBE showed, among various outcomes, a significant dose effect for Leydig cell tumors (LCTs), which are incidental in male Sprague-Dawley rats (Belpoggi *et al.* 1995; Belpoggi *et al.* 1997). For more detailed description of the experiment we refer the reader to the two papers by Belpoggi *et al.* (Belpoggi *et al.* 1995; Belpoggi *et al.* 1997). Upon a recent reanalysis by pathologists, some of the tumor pathologies were changed, and the new findings were published (Belpoggi *et al.* 1998). In the original paper, where three dose groups with $n = 60$ animals per group were studied (Belpoggi *et al.* 1995), the Hoel-Walburg method had been used and the findings were that the Leydig tumors were significantly increased in the high experimental groups over the control group. This analysis is also in agreement with the Dinse-Lagakos score test and the poly-3 test. A statistical analysis

using the revised pathology of the Leydig tumors using the poly-3 test, based on estimated time of death, was recently conducted and suggested that there was no dose effect on the Leydig cell tumors (Goodman *et al.* 2007). The authors of the re-analysis did not have the original data at the time and estimated the times of death for the individual animals. In contrast to the typical 2-year NTP study, the MTBE study was more than 3 years in duration and the last observed animal death was at 174 weeks. The Kaplan-Meier survival plots for these data are depicted in Figure 1.

We have analyzed this revised data first using the Dinse-Lagakos test (logistic regression score test) for incidental tumors, obtaining a two-sided p-value of 0.1. For the poly-k test, shown in Table 1 are the p-values as a function of the choice of k used in the risk adjustment for the animals. We also report p-values from some possible re-weighting schemes for the poly-k test. The reported p-values are from a two-sided test whereas Goodman *et al.* (2007) used a one-sided test. The disturbing issue here is that for the poly-k test we observed p-values ranging from $p = 0.03$ for $k = 1.5$ to an insignificant $p = 0.42$ for $k = 6$. The default value of $k = 3$ gives approximately the same p-value of 0.1 as did the logistic score test.

It should also be mentioned that if in this example the poly-k weights were applied up to 2 years (730 days) of age with a denominator of $t_{\max} = 730$ days, and then a full weight of 1 was used for animals dying after 2 years of age, which could be considered as one type of re-weighted poly-k approach, the range of p-values was more consistently between 0.01 and 0.02 for the various values of k (see polykw1 entries in Table 1). Here, it is important to point out that the estimated value of k for the time-to-death in the MTBE study obtained by fitting a Weibull model is about 7 (it is about 10 for the control dose group). Alternatively, one could use the usual poly-k weights up to 2 years (t_{\max} is observed maximum t) and then use a weight of 1 after 2 years (see polykw3 entries in Table 1). This leads to p-values ranging from 0.03 for $k = 1.5$ to 0.08 for $k = 6$. Another approach is to use $t_{\max} =$ the 95 percentile of the time-to-death distribution and apply a re-weighted poly-k. This results in p-values ranging from 0.02 for $k = 1.5$ to 0.19 for $k = 6$. Additional analysis results are reported in Table 1 for situations where outlier time points were removed. Further discussion is provided in the discussion section on the effect of different weighting schemes for adjusting the effect of differential survival.

The issue, however, from this practical example is: Is the poly-k test and especially poly-3 test truly robust and is it appropriate for lifetime rodent studies? How does its performance in terms of power compare with other tests? The poly-k test traditionally was applied to terminal sacrifice studies. The simulations supporting its use were from such terminal sacrifice studies and thus the robustness for lifetime studies needs to be established. For example, the distribution of survival times may become a critically important assumption if the poly-k test is applied to lifetime studies. This led us to carry out some simulation studies of the competing tests for either lethal or incidental tumors.

METHODS

Suppose the time-to-tumor onset and time-to-death without the tumor of interest are denoted by T_1 and T_2 , respectively. Let the time to the first of the two events be denoted by T and let $Y(t) = I(T_1 < T_2 \mid T = t)$ be an indicator of whether the tumor of interest is present at time t or not. With out loss of generality, let us assume that T is continuous. Define the hazard functions corresponding to tumor incidence and tumor-free death as $h_1(t)$ and $h_2(t)$, respectively, and the expected proportion of tumors that develop during the study by π . For a lifetime animal study, this proportion can be expressed as,

$$\pi = \int_0^{t_{\max}} h_1(u) \exp \left[p - \int_0^u (h_1(w) + h_2(w)) dw \right] du,$$

where t_{\max} is the death time of the last surviving animal in a lifetime study.

Now consider an experiment with n animals, where n_i of them are exposed to dose x_i ($i = 1, \dots, I$) over an extended period of time and n_0 animals are concurrently followed as controls ($x_0 = 0$). We assume that $x_0 < x_1 < \dots < x_I$. Let d_{ij} be the number of deaths with tumor in dose group i at time t_j in $[0, t_{\max}]$. Then, the number of events at the j th time of death with tumor is

$d_j = \sum_{i=0}^I d_{ij}$ and giving the total number of animals with the tumor to be d . Our interest is to test for a trend in tumor incidence rates (or presence of treatment effect) and there have been several approaches proposed to date. We will consider methods that do and do not adjust for differential mortality. Clearly, since the π can be confounded by differential mortality in each dose group, methods that ignore differential mortality could lead to biased conclusions. We briefly describe some of the methods that have been used in testing trend in tumor incidence.

We start from the most widely used trend test, the Cochran-Armitage test (Cochran 1954;

Armitage 1955). Let d_i be the number of animals with tumor in group i and $d = \sum_{i=0}^I d_i$. Under the null hypothesis (H_0) of no treatment effect on tumor onset, the expected number of animals with tumor in the i th group is $e_i = dp_i$, where $p_i = n_i/n$. The Cochran-Armitage linear trend test for proportions (Baker *et al.* 2007; Cochran 1954, Armitage 1955) is derived by pooling the entire duration of a study for each group and applying the test statistic given below for testing trend in proportions (which also corresponds to the score test statistic for testing $\beta = 0$ under a particular logistic model or the extended Mantel-Haenszel test (see Piegorsch and Bailer 1997, pp:228-234)

$$Z = \frac{\sum_{i=0}^I (d_i - e_i) x_i}{s},$$

where $s^2 = \frac{d(n-d)}{n(n-1)} \sum_{i=0}^I n_i (x_i - \bar{x})^2$, and \bar{x} is the mean dose. Under the null hypothesis H_0 , the test statistic Z has a standard normal asymptotic distribution. Clearly, this test does not adjust for any survival differences between the dose groups and is shown to be sensitive to increases in death due to high dose toxicity. Consequently the test may not control the type-I error rate (Bailer and Portier 1988).

The poly-k test (Bailer and Portier 1988; Portier and Bailer 1989) is derived as a direct modification of the Cochran-Armitage test by applying a weighting scheme for the contribution towards the denominator of the Cochran-Armitage test to be a power of the fraction of the time the animal survives. While all animals that survive to the end of the study irrespective of their tumor status or those who died early and have the tumor contribute fully with $w_{ij} = 1$ (weight for animal j in the i th dose group), those that die early without the tumor are assigned weights that are functions of their age defined by $w_{ij} = (t_{ij}/t_{\max})^k$ (where t_{\max} is the death time of last surviving animal in a lifetime study, t_{ij} is the observed death time for animal j in dose group i and k is non-negative). Then, the number at risk for the i th dose group is defined as the sum of these weights. It has been shown that for most tumors $k = 3$ leads to a robust test for trend (Portier and Bailer 1989). For details of the tests and the suggested variance correction see

Bieler and Williams (1993) and also Piegorsch and Bailer (1997). This test of trend for the study design employed by the NTP, *i.e.*, 2-year terminal sacrifice bioassay, has been endorsed as a preferred test with $k = 3$ as a default choice of k when it is not possible to decide whether the tumor under study is known to be lethal or non-lethal.

The logistic regression score test is obtained from the logistic regression model of tumor prevalence as a function of dose and survival time (linear treatment of dose and time) and inclusion of time as predictor adjusts for survival differences (Dinse and Lagakos 1983). Let the logistic model be defined as $\text{logit}(p) = \gamma_0 + \gamma_1 x + \gamma_2 t$, where p is the probability that the tumor is present at time t . The test statistic is the logistic score function for γ_1 divided by the (γ_1, γ_1) -element of the inverse of the observed information matrix, evaluated at $\gamma_1 = 0$ and the maximum likelihood estimates of γ_0 and γ_2 under the restriction that $\gamma_1 = 0$. Under the null hypothesis H_0 , the score test statistic is asymptotically normally distributed. For details of the score test see Dinse and Lagakos (1983).

Finally, the log-rank test (Kalbfleisch and Prentice 2002) is obtained by summing terms as in the Cochran-Armitage test on each stratum defined by each tumor death time and treating animals that die without the tumor as censored. Suppose there are $j = 1, \dots, J$ time intervals. Let n_{ij} denote the number of animals at risk in the j th stratum (stratum total equals $n_{.j}$) and i th group and let d_{ij} denote the number of animals “dying” with tumor in the j th stratum (stratum total equals $d_{.j}$) and i th group. Under the null hypothesis of no treatment effect on tumor mortality, the expected number of animals with tumor in the i th group and j th time is $e_{ij} = d_{.j}p_{ij}$, where $p_{ij} = n_{ij}/n_{.j}$. Similar to the poly- k test, the log-rank test is computed by applying the Mantel-Haenzel test statistic, which can be given by,

$$Z = \frac{\sum_{i=0}^I (d_i - e_i) x_i}{s},$$

where $s^2 = \sum_{j=1}^J \frac{d_{.j}(n_{.j} - d_{.j})}{n_{.j}(n_{.j} - 1)} \sum_{i=0}^I n_{ij} (x_i - \bar{x})^2$, and \bar{x} is the mean dose. Under null hypothesis H_0 , the test statistic Z has an asymptotic normal distribution. This method also accommodates differing follow up periods.

SIMULATION STUDY

The main goal of this study is to evaluate the robustness of the poly- k method, in particular when $k = 3$, using simulated data for a lifetime study without sacrifice, and to compare its performance with the log-rank test, which is valid for lethal tumors, and the logistic score test, which is valid for incidental tumors. We will also make comparisons with the Cochran-Armitage test, which does not account for differential survival but is valid when there is not differential survival between the groups as our reference. The simulation study will be helpful to understand the performance of the poly- k test (especially for $k = 3$) when lethality is assumed and when it is not. We considered two simulation scenarios. In both scenarios, the data were simulated to be similar on all characteristics except on the spacing of the dose levels and the effect of dose on mortality and tumor onset. Incidental tumors were simulated such that they can be observed only if the animal died due to causes other than the tumor of interest, whereas lethal tumors were simulated such that mortality was directly caused by the tumor under study. The observed outcome in each animal was determined using a Weibull model with hazard function given by

$$h(t_k|x) = \lambda_k \eta_k t_k^{\eta_k - 1} \exp(\beta_k x),$$

where λ_k is the scale parameter, η_k is the shape parameter and β_k is the log-hazard ratio for dose x and $k = 1, 2$.

The shape parameters η_k ($k = 1, 2$) decide the steepness of the tumor incidence and mortality, respectively, independent of the dose. The scale parameters λ_1 and λ_2 multiplied by the log-linear term are related to the dose and would determine the average time-to-tumor onset and time-to-death from competing risks in the control group ($X = 0$), respectively. The data on time to event (t) and event status (Y) for each animal were determined using the following procedure. For incidental tumors, the time-to-non-lethal tumor onset (t_1) and time-to-death (t_2) were computed, then the indicator of tumor presence at the time of death (Y) was assigned the value of 1 if $t_1 \leq t_2$ and time-at-death was $t = t_2$. For lethal tumors, the time-to-death from the tumor of interest (t_1) and time-to-death from other cause (t_2) were computed and the tumor was considered cause of death ($Y = 1$) if $t_1 \leq t_2$ and the time-to-death $t = \min\{t_1, t_2\}$. The two times t_1 and t_2 were generated independently.

In the first scenario we simulated data from four dose groups $X = 0$, $X = 0.5$, $X = 1$, and $X = 2$. The simulation study mimics Peddada *et al.* (Peddada *et al.* 2005) with a slight modification of the simulation parameters. The mortality shape parameter was fixed at $\eta_2 = 5$ and the corresponding scale parameter $\lambda_2 = 4.48 \times 10^{-8}$ so that control survival at 24 months is $\exp(-\lambda_2 t^{\eta_2}) = \exp(-4.48 \times 10^{-8} \times 24^5) = 0.7$ or $\exp(-4.48 \times 10^{-8} \times (24/730)^5 \times 24^5) = 0.7$ when $t = 730$ days. Three different values of dose effect on competing risk rates [$\exp(\beta_2) = 1.0, 1.5$ and 2.0] ranging from no dose effect on mortality to high dose effect were considered. For tumor onset, we studied three shape parameters 1.5, 3 and 6 and six scale parameters corresponding to rare tumor rate ($\pi = 0.05$) and common tumor rate ($\pi = 0.30$) in the control group ($X = 0$) for each shape parameter. In Table 3 are given the parameters used in each simulation study, values of dose effect on tumor onset [$\exp(\beta_1) = 1.0, 1.5$ and 2.0] and the probabilities of $Y = 1$ (having tumor in the incidental case or death from tumor in the mortality case) for some of the studied conditions. For instance, when $\eta_1 = 1.5$ and $\lambda_1 = 0.0025$ (a common tumor scenario), shown in Table 3 are the tumor rates for $\beta_1 = \beta_2 = 0$ to be 0.3 while for $\beta_1 = 0.42$ and $\beta_2 = 0$ the rates increase from 0.30 to 0.52. Tumor rates for rare tumors were also computed for different scenarios, *e.g.*, when $\eta_1 = 1.5$, $\lambda_1 = 0.00038$ and $\beta_1 = \beta_2 = 0$, the rate was 0.05 while for $\beta_1 = 0.42$ and $\beta_2 = 0$ the rates increased from 0.05 to 0.11 (table not reported).

In the second scenario we simulated data with three dose-groups in which there were 100 animals in each group. The doses were evenly spaced with $X = 0$ (control), $X = 1$ (low dose) and $X = 2$ (high dose). In simulating the data, a Weibull tumor onset distribution with two shape parameters, 4.2 for the censored and 6.4 for those that died with tumor and two scale parameters 0.001010 and 0.001205 taken from fitting a Weibull model to liver carcinoma data from a B6C3F₁ mouse study were used. This was done by considering seven different dose effects on tumor onset or tumor lethality [$\exp(\beta_1) = 0.6, 1.0, 1.4, 1.6, 1.8, 2.0$, and 2.2] and three dose effects on competing risk rates [$\exp(\beta_2) = 0.6, 1.0$, and 1.4] that lead to seven different overall tumor onset probabilities and three different competing risk rates.

The type-I error rate and power evaluations for a two-sided test were based on a nominal significance level of 5%. Two thousand simulated datasets for each of the different parameter configurations were generated and these would have a margin of error about 1%. Analysis was then made using the Cochran-Armitage, logistic regression, poly-k ($k = 1.5, 2.0, 3.0, 4.0, 5.0$, and 6.0) and log-rank tests of trend. We also considered some potential re-weighting schemes for the poly-k test (mentioned in the introduction and discussion sections) that use different

weighting schemes for tumor free deaths and hence the contribution of these animals to the denominator of the test statistic.

RESULTS

Several scenarios were examined both for the situation where the tumor is lethal and is declared a cause of death and the case where the tumor is incidental and was not considered to be related to the cause of death. To begin, consider the simplest situation where cancer is increased with increasing dose while the other causes of death are not affected by the exposure. The statistical power for increasing tumor hazard rates for the case where the tumor is instantly lethal and is a cause of death is shown in Panel A of Figure 2. We see that the tests all behave properly; however, the log-rank test, which necessarily assumes that the tumor is the cause of death has a power much greater than that of the poly-k, consistent with the fact that log-rank is the most powerful test for fatal tumors while the logistic regression score test is most powerful for incidental tumors. The Cochran-Armitage test performed similarly to the poly-1.5 test. The poly-k tests were more conservative with lower type-I error rates (rejects less frequently than nominally specified under the null hypothesis or observed $\alpha < \text{nominal } \alpha$) that decreased with increasing value of k (see Table 4). In Panel B of Figure 2 is presented a similar situation for the case where the tumors are incidental except now the logistic regression analysis replaces the log-rank test for the incidental case. Again, the test, which assumes knowledge of the cause of death has power considerably greater than that of the poly-k test as well as the Cochran-Armitage test, which was similar to the poly-1.5 test but superior to the poly-3 and poly-6 tests. Similar results are reported in Table 5 for common tumors and Table 6 for rare tumors.

We next considered the opposite scenario where the chemical affects the cause of death for the non-tumor mortality but has no effect on the tumors themselves. The case where tumors are considered to be a cause of death is shown in Panel C of Figure 2. What we observe is that the log-rank test as well as the poly-6 test behave properly, that is, both of them had the expected power of rejecting the null hypothesis regardless of dose effect on the non-cancer causes of death. The other two poly-k tests and the Cochran-Armitage test all proceed to have increasing probabilities of rejecting the null hypothesis when in fact it should be accepted since the chemical is not affecting the tumor. The corresponding situation for the case where tumors are incidental is presented in Panel D of Figure 2. The situation is similar as with Panel C of Figure 2. In both situations the poly-1.5 and the poly-3 test as well as the Cochran Armitage test give misleading conclusions concerning the statistical significance of tumorigenicity when in fact there is no dose effect on the incidence of tumors.

The type-I error rates corresponding to the different tests for rare as well as common tumor scenarios are given in Table 4. As expected the logrank test for lethal tumors and the logit score test for incidental tumors maintained the nominal type-I error rate irrespective of whether treatment dose affects lethality or not. On the other hand the poly-3 test was liberal, rejecting more than the expected 95% when k was 3 or less, but was conservative when k was larger than 3. For instance, the type-I error rate increased from 0.019 to 0.265 when k changed from $k = 3$ to $k = 6$. This effect was less pronounced when the tumor was rare.

The results of the scenario where the chemical affects the tumors but it also affects the non-cancer mortality are shown in Panels A and B of Figure 3. The two figures show the probability of rejecting the null hypothesis of the statistical test as the tumor hazard ratio is increased. What is observed is that both the log-rank test and the logistic test behave properly with good power in their corresponding situations of lethal tumor and incidental tumor scenarios, respectively, when compared to the other tests. The poly-k test for $k = 1.5$ and 3 as well as the Cochran-Armitage test all incorrectly reject the null hypothesis when the tumor is not affected by the

exposure, *i.e.*, hazard-ratio equal to 1.0. What these tests are saying is not that there is a positive effect of the chemical on the tumor but that in fact there is a protective effect that is of course incorrect in this situation. Also the power of both the log-rank test and of the logistic test is considerably greater than that of the poly-k and the Cochran-Armitage tests. When we considered similar situations except now the chemical has a protective effect for non-cancer mortality (data not shown) in other words at higher doses the animals tend to live longer; which is a very uncommon situation but it is what apparently occurred in the previously mentioned analysis of MTBE data, we observed the same problem with the tests. Other than the log-rank test and the logistic test, all other tests were not behaving properly for the situation where the chemical was not affecting the tumor.

Finally we consider the remaining scenario of the exposure affecting both the tumors and the non-tumor mortality. The effect on the tumor hazard ratio is set at 2.2 and the hazard ratio for non-tumor mortality is varied from no effect at 1.0 up to 2.2. What is observed in Panels C and D of Figures 3 as well as Tables 5 and 6 is that the poly-k and Cochran Armitage test rapidly lose power compared to the log-rank and logistic tests with increasing effects on the non-cancer mortality. As shown in Table 5, when the true shape parameter changes from $k = 3$ to $k = 6$ the power of the poly-3 test decreases from 0.98 to 0.88 in the case of lethal tumors and from 0.96 to 0.78 in incidental tumors. This problem is minimal when the tumor is rare.

DISCUSSION

The application of the poly-k test in lifetime animal cancer studies was examined and was shown to have lower power compared to the log-rank and logistic tests since the latter two tests incorporate lethality information. In the 2-year sacrifice studies as used by the NTP the choice of test and power differences should not matter nearly as much since typically most animals will be sacrificed. The problem in lifetime studies besides the loss of power is that if the chemical simply affects general survival the poly-k test may incorrectly conclude that the chemical also affects the occurrence of tumors and is thus an animal carcinogen. This is clearly an unacceptable situation. We conclude that lethality information and the use of the appropriate statistical test are necessary in lifetime studies especially when differential survival is present.

In addition to the scenarios that are reported in the graphs we also studied tumor rate scenario between 5% and 30% and the results were consistent and similar to the rare and common tumor scenarios and hence are not reported. We also studied several other methods such as the Hoel-Walburg test, Gart's trend test (Gart *et al.* 1979), and the Jonckheer-Terpstra test (Hollander and Wolfe 1999) for incidental tumors and they performed no better than the reported methods. Similarly, we also considered other variations of poly-k ($k = 2, 4, \text{ and } 5$) and the results were similar to those reported.

In a limited simulation study we studied the effect of several different weighting schemes for adjusting differential survival and found the results to be interesting and warrant further investigation. Accordingly, we are studying these methods in further detail both analytically and in a simulation study. Our limited simulations showed that these weighting schemes could lead to more robust survival adjusted tests than the poly-k test that uses the usual weights. The poly-k weight for an animal dying at time t_i is $(t_i/t_{\max})^k$, where t_{\max} is the largest death time observed in the study. In a typical 2-year study, many animals die at t_{\max} as a result of the terminal sacrifice. In a lifetime study, however, typically a single animal dies at t_{\max} . Thus, in contrast to a 2-year study, a single animal in a lifetime study might exert undue influence. As an extreme example, suppose that most animals die before 2.5 years, but one animal survives for 4 years. The single animal dying at 4 years will cause all other (tumor-free) animals to have small weights, especially for large values of k , thus reducing the effective sample size and exaggerating the tumor response relative to an otherwise identical data set without this one

animal. One possibility is that the poly-k test performs poorly in lifetime studies for which the death time determining t_{\max} is an outlier, but performs reasonably well otherwise. One strategy that is used in some life studies is to sacrifice the animals in a dose group after 95% of them have died. We could apply the same idea and set t_{\max} at the 95% percentile in which avoiding the one or two old age outliers. We could also use arguments similar to ones used in regression analysis to remove outlier time points and reanalyze the data using the poly-k method. This example illustrates the need and importance of investigating the pattern of deaths in the analysis of animal bioassay data.

Our simulation results in general show that the poly-k test is not robust for testing trend in some lifetime animal studies, especially the poly-3 test was not at all robust (did not result in similar p-values for data generated from different shape parameter values). So, with lifetime studies that result in differential survival the problem remains what analysis to use when the pathologist is not able to decide whether the tumor is incidental or a cause of death. We have shown that the unmodified poly-k test is not a solution to the issue and can in some circumstances produce a misleading result. Peto *et al.* (STP Peto Working Group 2002) have provided methods in which the degree of lethality is included in the analysis. We do not have a solution and do not fully know at this time how serious this issue is. Currently, we are conducting simulations for various scenarios using modified poly-k tests based on interesting observations we have from application to the data example. The results are reported in Tables 1 and 2. Both tables show that while the poly-k test exhibits inconsistency or non-robustness to changes in the shape parameter k , some modification of the weighting scheme leads to a more consistent and robust results.

The proposed modifications of the poly-k (re-weighted poly-k) would address issues of outliers and they are less sensitive to lethality determination by pathologists. But they need further investigation as what we have so far is their application to the MTBE data example. We are studying these further using simulation studies. The way the re-weighting works is it avoids the problem from the application of the poly-k weights (Bailer and Portier 1988; Portier and Bailer 1989), which would heavily reduce effective sample size since animals that die tumor free early on in the study would get very small weights and hence would contribute less to the denominator of the test statistic.

The proposed re-weighting schemes are:

- (i) using a weight with t_{\max} = the 95 percentile of the time-to-death distribution as the denominator for the poly-k weights and weight animals that die tumor free before the end of the study accordingly up to t_{\max} but give full weight to those who survived beyond t_{\max} and to those that die with the tumor
- (ii) using t_{\max} = 2 years (730 days) as the denominator for the poly-k weights and weighting animals that die tumor free before the end of the study accordingly up to 2 years but giving full weight to those who survived beyond two years and those that die with the tumor
- (iii) using t_{\max} = maximum observed time-to-death as the denominator for the poly-k weights and weight animals that die tumor free accordingly up to 2 years but give full weight to those who survived beyond two years and those that die with the tumor
- (iv) calculating t_{\max} after removing animals with observed outlier time-to-death and using either of the above three suggested weighting schemes on the remaining animals.
- (v) Do the above three re-weighting schemes after totally removing animals with observed outlier time-to-death. The removing of outliers could be made on similar arguments used in regression analysis.

Another approach for sacrifice data is to use an estimated k (Moon *et al.* 2003), which is in line with the Bailer and Portier (1988) observation that if the shape of the tumor incidence functions is expected to follow time to some power k , poly- k would have superior operating characteristics. We plan to explore this technique, the order restricted test in (Peddada *et al.* 2005) and our proposed weighting schemes in our future work. We also plan to investigate the robustness of these methods to the assumption of a Weibull distribution made on time-to-death data.

Acknowledgments

This work was partially supported by NSF (EPS-0447660, NSF/EPSCoR 2004 RII), and MUSC office of the provost.

REFERENCES

- Armitage P. Tests for linear trends in proportions and frequencies. *Biometrics* 1955;11:375–86.
- Bailer AJ, Portier CJ. Effects of treatment-induced mortality and tumor-induced mortality on tests for carcinogenicity in small samples. *Biometrics* 1988;44:417–31. [PubMed: 3390507]
- Baker GS, Nakamura DW, Hoel DG. Comparison of two models of cancer risk estimation: a statistical analysis. *European J Oncology* 2007;11:165–76.
- Belpoggi F, Soffritti M, Maltoni C. Methyl-tertiary-butyl ether (MTBE)-a gasoline additive-causes testicular and lymphohaematopoietic cancers in rats. *Toxicol Indust Health* 1995;11:119–49.
- Belpoggi F, Soffritti M, Filippini F, et al. Results of long-term experimental studies on the carcinogenicity of methyl tert-butyl ether. *Annals NY Acad Sci* 1997;837:77–95.
- Belpoggi F, Soffritti M, Maltoni C. Pathological characterization of testicular tumors and lymphomas-leukemias, and of their precursors observed in Spague-Dawley rats exposed to methyl-tertiary-butyl-ether (MTBE). *European J Oncology* 1998;3:201–6.
- Belpoggi F, Soffritti M, Padovani M, et al. Results of long-term carcinogenicity bioassay on Sprague-Dawley rats exposed to aspartame administered in feed. *Annals NY Acad Sci* 2006;1076:559–77.
- Bieler G, Williams R. Ratio estimates, the delta method and quantal response tests for increased carcinogenicity. *Biometrics* 1993;49:793–801. [PubMed: 8241374]
- Cochran WG. Some methods of strengthening the common Chi-square tests. *Biometrics* 1954;10:417–51.
- Dinse GE. A comparison of tumor incidence analyses applicable in single-sacrifice animal experiments. *Statistics in Medicine* 1994;13:689–708. [PubMed: 8023043]
- Dinse GE, Lagakos SW. Regression analysis of tumor prevalence data. *App Statistics* 1983;32:236–48.
- Gart J, Chu K, Tarone R. Statistical Issues in the interpretation of chronic bioassay tests for carcinogenicity. *J NCI* 1979;62:957–74.
- Goodman JE, Gaylor DW, Beyer LA, et al. Effects of MTBE on Leydig cell tumors in Sprague–Dawley rats: range of possible poly-3 results. *Regul Toxicol Pharmacol.* 2007 accepted for publication. doi: 10.1016/j.yrtph.2007.05.002.
- Hoel DG, Walburg HEJ. Statistical analysis of survival experiments. *J NCI* 1972;49:361–72.
- Hollander, M.; Wolfe, DA. *Nonparametric Statistical Methods.* Wiley & Sons Inc; Weinheim, Germany: 1999.
- Kalbflied, JD.; Prentice, RL. *The Statistical Analysis of Failure Time Data.* John Wiley and Sons; Hoboken, NJ, USA: 2002.
- Moon H, Ahn H, Kodell RL, et al. Estimation of k for the poly- k test with application to animal carcinogenicity studies. *Statistics in Medicine* 2003;22:2619–36. [PubMed: 12898548]
- Peddada DS, Dinse GE, Haseman JK. A survival adjusted quantal response test for comparing tumor incidence rates. *Appl Statistics* 2005;54:51–61.
- Peto, R.; Pike, MC.; Day, NE., et al. *IARC Monographs, Supplement 2: Long-term and Short-term Screening Assays for Carcinogens: A Critical Appraisal.* International Agency for Research on Cancer; Lyon, France: 1980. Guidelines for Simple, Sensitive Significance Tests for Carcinogenic Effects in Long-Term Animal Experiments; p. 311–426.

- Piegorsch, WW.; Bailer, AJ. *Statistics for Environmental Biology and Toxicology*. Chapman & Hall; New York, NY, USA: 1997.
- Portier CJ, Bailer AJ. Testing for increased carcinogenicity using a survival-adjusted quantal response test. *Fundamental App Toxicol* 1989;12:731–7.
- Portier CJ, Hedges JC, Hoel DG. Age-specific models of mortality and tumor onset for historical control animals in the National Toxicology Program's carcinogenicity experiments. *Cancer Res* 1986;46:4372–8. [PubMed: 3731095]
- STP Peto Working Group. Statistical methods for carcinogenicity studies. *Toxicologic Pathology* 2002;30:403–12. [PubMed: 12051559]

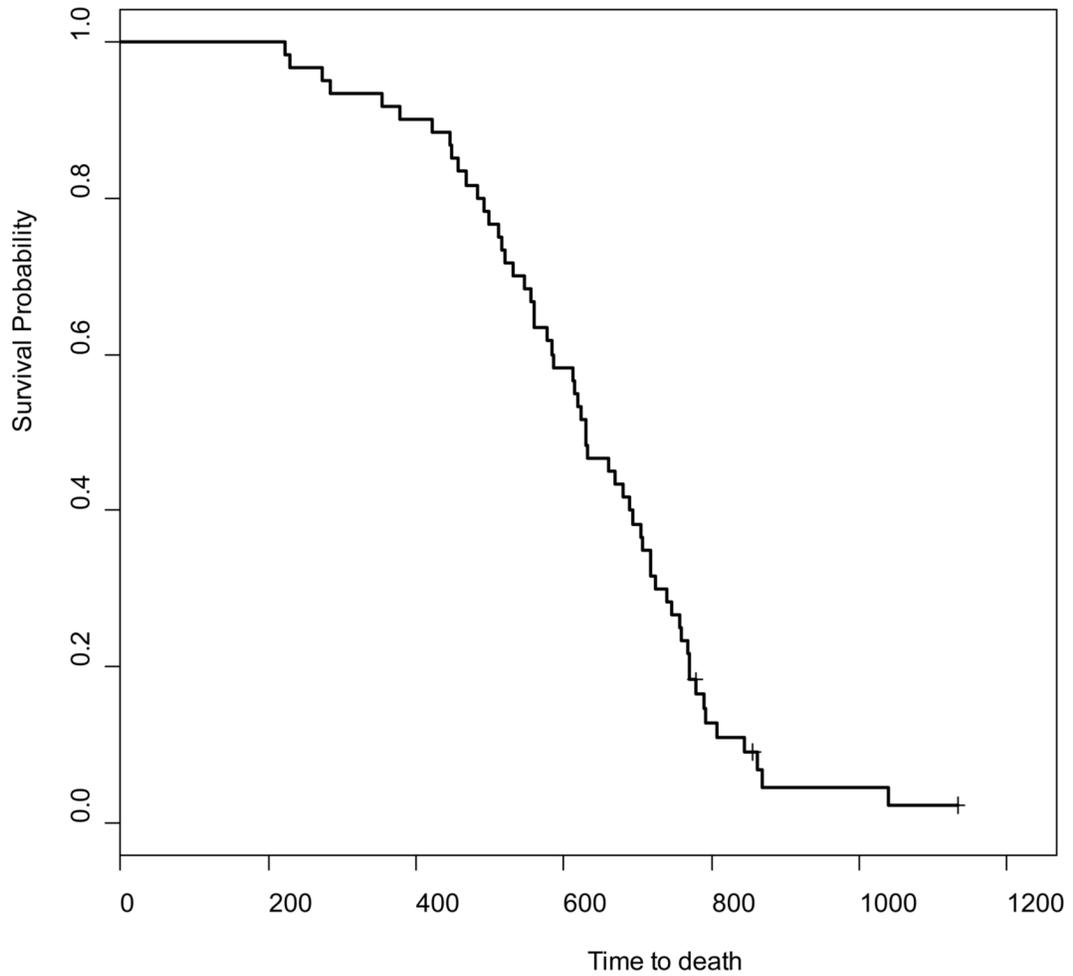


Figure 1. Kaplan-Meier Survival Curves for the three dose groups in the Leydig tumors MTBE data example (solid line = control dose, broken line = 250mg/kg/day, dotted line = 1000mg/kg/day). There were 3/60, 5/60 and 11/60 LCT events for control, 250mg/kg and 1000mg/kg doses, respectively, and are indicated by + signs in the plots.

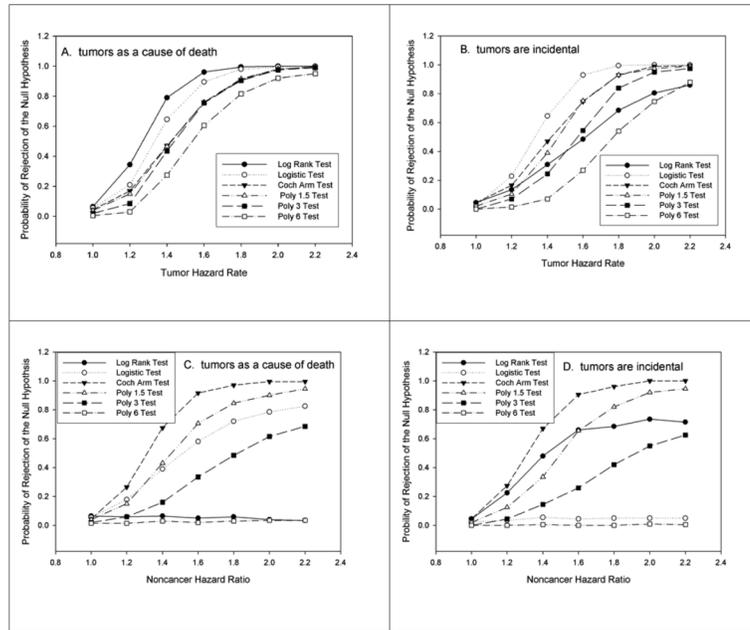


Figure 2. Test size and power of trend tests for equally spaced three dose groups and $n = 100$ per group (shape and scale parameters for other causes of death were fixed at 4.2 and 0.01205, respectively). Panel 2A and 2B are when treatment does not affect death due to causes other than the tumor of interest, *i.e.*, non-cancer hazard ratio $HR = 1$ and Panel 2C and 2D are when treatment does not affect the tumor of interest, *i.e.*, tumor onset hazard rate = 1 but affects non-cancer mortality.

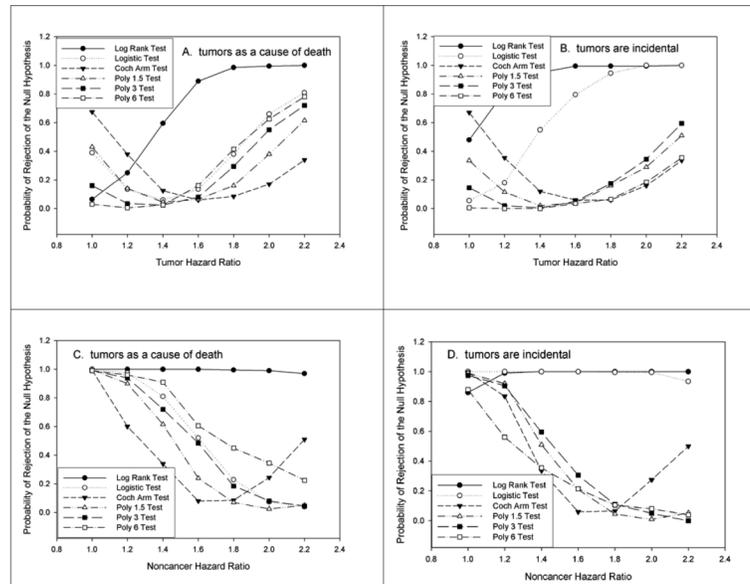


Figure 3. Test size and power of trend tests for equally spaced three dose groups and $n = 100$ per group (shape and scale for other causes of death were fixed at 4.2 and 0.01205, respectively). Panel 3A and 3B are when the treatment effect on death due to other causes is fixed at hazard ratio $HR = 1.4$ and Panel 3C and 3D are when the treatment effect on tumor incidence is fixed at $HR = 2.2$ and when the treatment affects non-cancer mortality.

Effect on p-value of different poly-k weighting schemes as applied to the Leydig tumors MTBE data example with three dose groups and 60 animals per group. P-values are two sided. The p-value from a logistic score test was 0.09 and it was 0.11 after removing animal at t = 1218 days.

Table 1

Method	Analysis including all animals with different values of tmax					
	k = 1.5	k = 2.0	k = 3.0	k = 4.0	k = 5.0	k = 6.0
Polyk	0.0298	0.0429	0.0894	0.1704	0.2852	0.4200
Polykw1	0.0121	0.0122	0.0128	0.0140	0.0154	0.0171
Polykw2	0.0220	0.0282	0.0478	0.0797	0.1256	0.1850
Polykw3	0.0270	0.0346	0.0512	0.0653	0.0752	0.0812
	Analysis including all animals but tmax = 1133 days instead of tmax = 1218 days					
Polyk	0.0275	0.0385	0.0762	0.1415	0.2364	0.3567
	Analysis after removing the outliers at t = 1133 and t = 1218 days (tmax = 976 days)					
polyk	0.0204	0.0267	0.0476	0.0834	0.1375	0.2105
	Analysis after removing the outlier animals at t = 1218 days					
Polyk	0.0463	0.0616	0.1112	0.1910	0.2993	0.4262
Polykw1	0.0229	0.0229	0.0239	0.0257	0.0279	0.0304
Polykw2	0.0385	0.0476	0.0755	0.1183	0.1766	0.2488

polyk: p-value from a poly-k test using tmax = 1218 days as denominator for all non tumor deaths

polykw1: p-value from a poly-k test using tmax = 2 years (730 days) as denominator

polykw2: p-value from a re-weighted poly-k test using tmax = 956 days (95% percentile of t) as denominator

polykw3: p-value from a re-weighted poly-k test using tmax = 1218 as denominator for t < 2 years or 730 days and full weight from two years

Table 2

Effect on tumor prevalence rate of different poly-k weighting schemes applied to the Leydig tumors MTBE data example with three dose groups and 60 animals per group and k = 1.5, 3.0 and 6.0. The observed tumor prevalence rates are, p1 = 0.050, p2 = 0.0833 and p3 = 0.1833 for dose groups = 0mg/kg/day, 250mg/kg/day and 1000mg/kg/day.

Analysis including all animals with different values of tmax									
prevalence rate	k = 1.5			k = 3.0			k = 6.0		
	p1	p2	p3	p1	p2	p3	p1	p2	p3
Polyk	0.125	0.193	0.367	0.253	0.350	0.543	0.591	0.679	0.776
Polykw1	0.066	0.107	0.238	0.078	0.124	0.274	0.096	0.148	0.314
Polykw2	0.091	0.142	0.290	0.143	0.206	0.375	0.267	0.330	0.483
Polykw3	0.097	0.150	0.297	0.134	0.193	0.351	0.163	0.222	0.376
Analysis including all animals but tmax = 1133 days instead of tmax = 1218 days									
Polyk	0.114	0.176	0.342	0.215	0.302	0.489	0.484	0.578	0.691
Analysis after removing the outliers at t = 1133 and t = 1218 days (tmax = 976 days)									
polyk	0.065	0.291	0.274	0.104	0.216	0.357	0.203	0.359	0.454
Analysis after removing the outlier animals at t = 1218 days									
Polyk	0.114	0.176	0.321	0.215	0.302	0.465	0.484	0.578	0.671
Polykw1	0.056	0.107	0.221	0.078	0.124	0.256	0.096	0.148	0.294
Polykw2	0.091	0.142	0.271	0.143	0.206	0.353	0.267	0.330	0.459

polyk: tumor prevalence rates from a poly-k test using tmax = 1218 days as denominator for all non tumor deaths

polykw1: tumor prevalence rates from a poly-k test using tmax = 2 years (730 days) as denominator

polykw2: tumor prevalence rates from a re-weighted poly-k test using tmax = 956 days (95% percentile of t) as denominator

polykw3: tumor prevalence rates from a re-weighted poly-k test using tmax = 1218 as denominator for t < 2 years or 730 days and full weight from two years

Table 3

Probability of having an event (death for incidental tumor and death due to tumor for lethal tumor) for selected parameters of overall tumor rate and overall mortality rate. Four dose groups with doubling spacing and $n = 50$ per group ($\eta_2 = 5.0$ and $\lambda_2 = 0.448 \times 10^{-7}$).

$(\eta_1 = 1.5 \text{ and } \lambda_1 = 0.0025, \exp(\beta_2) = 1.0)$			
Probability	$\exp(\beta_1) = 1.0$	$\exp(\beta_1) = 1.5$	$\exp(\beta_1) = 2.0$
$\Pr(Y = 1 X = 0)$	0.30	0.30	0.30
$\Pr(Y = 1 X = 0.5)$	0.30	0.35	0.39
$\Pr(Y = 1 X = 1.0)$	0.30	0.41	0.50
$\Pr(Y = 1 X = 2.0)$	0.30	0.52	0.74
$(\eta_1 = 1.5 \text{ and } \lambda_1 = 0.0025, \exp(\beta_2) = 2.0)$			
Probability	$\exp(\beta_1) = 1.0$	$\exp(\beta_1) = 1.5$	$\exp(\beta_1) = 2.0$
$\Pr(Y = 1 X = 0)$	0.30	0.30	0.30
$\Pr(Y = 1 X = 0.5)$	0.27	0.32	0.36
$\Pr(Y = 1 X = 1.0)$	0.25	0.35	0.43
$\Pr(Y = 1 X = 2.0)$	0.21	0.41	0.59
$(\eta_1 = 3.0 \text{ and } \lambda_1 = 0.000017, \exp(\beta_2) = 1.0)$			
Probability	$\exp(\beta_1) = 1.0$	$\exp(\beta_1) = 1.5$	$\exp(\beta_1) = 2.0$
$\Pr(Y = 1 X = 0)$	0.304	0.304	0.304
$\Pr(Y = 1 X = 0.5)$	0.304	0.354	0.393
$\Pr(Y = 1 X = 1.0)$	0.305	0.409	0.492
$\Pr(Y = 1 X = 2.0)$	0.305	0.528	0.700
$(\eta_1 = 3.0 \text{ and } \lambda_1 = 0.000017, \exp(\beta_2) = 2.0)$			
Probability	$\exp(\beta_1) = 1.0$	$\exp(\beta_1) = 1.5$	$\exp(\beta_1) = 2.0$
$\Pr(Y = 1 X = 0)$	0.304	0.304	0.304
$\Pr(Y = 1 X = 0.5)$	0.258	0.303	0.338
$\Pr(Y = 1 X = 1.0)$	0.219	0.302	0.374
$\Pr(Y = 1 X = 2.0)$	0.152	0.300	0.451

Table 4

Test size comparisons of trend tests on four doses (0,0.5,1.0 and 2.0) and n = 50 per group based on 2000 simulated data sets, where CA: Cochran-Armitage trend test, Logit: Dinse-Lagakos logistic regression score test, Poly-k: survival adjusted test by Bailer and Portier (1988); β_1 is effect of dose on tumor incidence or mortality due to tumor; β_2 is effect of dose on mortality due to competing risk or non-tumor related death.

Common Tumors									
Tumor	η_1	(β_1, β_2)	CA	logrank	logit	Poly-k			
						k = 1.5	k = 3	k = 6	
Cause of death	1.5	(0.0,0.0)	.042	.040	.047	.023	.010	.005	
		(0.0,0.7)	.174	.046	.779	.027	.030	.099	
	3.0	(0.0,0.0)	.051	.048	.050	.031	.017	.007	
		(0.0,0.7)	.456	.042	.809	.135	.019	.013	
	6.0	(0.0,0.0)	.048	.046	.049	.032	.018	.006	
		(0.0,0.7)	.857	.046	.649	.572	.265	.017	
Tumor									
Incidental	1.5	(0.0,0.0)	.042	.045	.040	.022	.009	.003	
		(0.0,0.7)	.174	.876	.043	.026	.026	.077	
	3.0	(0.0,0.0)	.051	.054	.050	.030	.013	.002	
		(0.0,0.7)	.456	.817	.043	.129	.015	.008	
	6.0	(0.0,0.0)	.048	.054	.046	.026	.009	.001	
		(0.0,0.7)	.857	.655	.045	.556	.220	.005	
Rare Tumors									
Tumor	η_1	(β_1, β_2)	CA	logrank	logit	Poly-k			
						k = 1.5	k = 3	k = 6	
Cause of death	1.5	(0.0,0.0)	.046	.049	.030	.044	.037	.028	
		(0.0,0.7)	.067	.049	.258	.044	.058	.141	
	3.0	(0.0,0.0)	.049	.051	.037	.044	.042	.032	
		(0.0,0.7)	.092	.051	.157	.047	.037	.067	
	6.0	(0.0,0.0)	.049	.050	.038	.045	.044	.030	
		(0.0,0.7)	.179	.047	.054	.081	.035	.022	

		Common Tumors									
Tumor											Poly-k
Tumor											Poly-k
Incidental	η_1	(β_1, β_2)	CA	logrank	logit	k = 1.5	k = 3	k = 6			
	1.5	(0.0,0.0)	.046	.053	.038	.044	.037	.027			
		(0.0,0.7)	.067	.337	.043	.044	.058	.136			
	3.0	(0.0,0.0)	.049	.053	.039	.044	.041	.031			
		(0.0,0.7)	.092	.260	.035	.043	.036	.065			
	6.0	(0.0,0.0)	.049	.049	.040	.045	.043	.029			
		(0.0,0.7)	.179	.189	.030	.081	.034	.021			

Table 5

Statistical power comparisons of trend tests on four doses (0,0.5,1.0 and 2.0) and n = 50 per group based on 2000 simulated data sets, where CA: Cochran-Armitage trend test, Logit: Dinse-Lagakos logistic regression score test, Poly-k: survival adjusted test by Bailer and Portier (1988); β_1 is effect of dose on tumor incidence or mortality due to tumor; β_2 is effect of dose on mortality due to competing risk or non-tumor related death (for common tumor).

Tumor	η_1	(β_1, β_2)	CA	logrank	logit	Poly-k			
						k = 1.5	k = 3	k = 6	
Cause of death	1.5	(0.4,0.0)	.749	.791	.366	.657	.522	.282	
			.998	1.00	.768	.995	.985	.862	
			.221	.695	.242	.457	.638	.734	
	3.0	(0.7,0.7)	.896	.997	.052	.975	.988	.975	
			.671	.795	.478	.617	.505	.275	
			.992	.999	.889	.988	.978	.863	
	6.0	(0.4,0.7)	.051	.600	.256	.087	.200	.314	
			.356	.984	.047	.641	.792	.816	
			.486	.754	.544	.452	.394	.234	
		(0.7,0.0)	.900	.999	.929	.903	.881	.736	
			.405	.440	.185	.099	.020	.017	
			.089	1.00	.085	.024	.040	.051	
Incidental	η_1	(β_1, β_2)	CA	logrank	logit	k = 1.5	k = 3	k = 6	
			1.5	.749	.512	.756	.641	.487	.188
				.998	.913	.999	.995	.979	.739
	3.0	(0.4,0.7)	.221	.999	.611	.446	.605	.609	
			.896	1.00	.986	.972	.983	.931	
			.671	.464	.743	.586	.436	.137	
		(0.7,0.0)	.992	.865	.998	.987	.958	.639	
			.051	.999	.534	.080	.173	.208	
			.356	1.00	.958	.621	.732	.630	
	6.0	(0.4,0.0)	.486	.353	.645	.410	.268	.044	
			.896	.671	.984	.878	.776	.280	

Tumor	η_1	(β_1, β_2)	CA	logrank	logit	Poly-k		
						k = 1.5	k = 3	k = 6
		(0.4, 0.7)	.401	.980	.376	.090	.013	.005
		(0.7, 0.7)	.089	.917	.044	.027	.070	.165

Table 6

Statistical power comparisons of trend tests on four doses (0,0.5,1.0 and 2.0) and n = 50 per group based on 2000 simulated data sets, where CA: Cochran-Armitage trend test, Logit: Dinse-Lagakos logistic regression score test, Poly-k: survival adjusted test by Bailer and Portier (1988); β_1 is effect of dose on tumor incidence or mortality due to tumor; β_2 is effect of dose on mortality due to competing risk or non-tumor related death (for rare tumor).

Tumor	η_1	(β_1, β_2)	CA	logrank	logit	Poly-k		
						k = 1.5	k = 3	k = 6
Cause of death	1.5	(0.4,0.0)	.246	.244	.163	.224	.195	.132
		(0.7,0.0)	.670	.672	.485	.646	.602	.466
		(0.4,0.7)	.091	.230	.111	.206	.340	.523
	3.0	(0.7,0.7)	.402	.650	.040	.610	.761	.871
		(0.4,0.0)	.239	.236	.204	.220	.196	.147
		(0.7,0.0)	.658	.673	.586	.637	.602	.475
	6.0	(0.4,0.7)	.049	.170	.070	.076	.133	.262
		(0.7,0.7)	.127	.465	.038	.255	.396	.571
		(0.4,0.0)	.215	.223	.207	.204	.186	.141
		(0.7,0.0)	.581	.650	.590	.578	.557	.472
		(0.4,0.7)	.088	.144	.034	.042	.033	.064
		(0.7,0.7)	.051	.314	.039	.051	.082	.166
Tumor								
Incidental	η_1	(β_1, β_2)	CA	logrank	logit	Poly-k		
						k = 1.5	k = 3	k = 6
	1.5	(0.4,0.0)	.246	.223	.223	.224	.194	.131
		(0.7,0.0)	.670	.603	.662	.645	.599	.452
		(0.4,0.7)	.091	.802	.192	.205	.338	.513
	3.0	(0.7,0.7)	.402	.975	.586	.610	.755	.857
		(0.4,0.0)	.239	.225	.221	.218	.195	.141
		(0.7,0.0)	.658	.578	.662	.633	.596	.454
		(0.4,0.7)	.049	.617	.137	.076	.131	.253
		(0.7,0.7)	.127	.890	.388	.254	.392	.558
		(0.4,0.0)	.215	.205	.217	.203	.183	.133
	(0.7,0.0)	.581	.534	.633	.573	.546	.434	

Tumor	η_1	(β_1, β_2)	CA	logrank	logit	Poly-k		
						k = 1.5	k = 3	k = 6
		(0.4,0.7)	.088	.422	.104	.041	.033	.062
		(0.7,0.7)	.051	.682	.238	.050	.082	.157