

# Estimation of $k$ for the Poly- $k$ Test

## with Application to Animal Carcinogenicity Studies

**Hojin Moon<sup>1</sup>, Hongshik Ahn<sup>2</sup>, J. Jack Lee<sup>1</sup> and Ralph L. Kodell<sup>3</sup>**

<sup>1</sup>Department of Biostatistics  
The University of Texas M. D. Anderson Cancer Center  
1515 Holcombe Boulevard - 447, Houston, TX 77030-4009  
*email:* hojin@odin.mdacc.tmc.edu

<sup>2</sup>Department of Applied Mathematics and Statistics  
State University of New York at Stony Brook  
Stony Brook, NY 11794-3600

<sup>3</sup>Division of Biometry and Risk Assessment  
National Center for Toxicological Research  
Food and Drug Administration  
Jefferson, AR 72079

### SUMMARY

This paper extends the survival-adjusted Cochran-Armitage test in order to achieve improved robustness to a variety of tumor onset distributions. The Cochran-Armitage test is routinely applied for detecting a linear trend in the incidence of a tumor of interest across dose groups. To improve the robustness to the effects of differential mortality across groups, Bailer and Portier (1988, *Biometrics*, **44**, 417-431) introduced the Poly-3 test by a survival adjustment using a fractional weighting scheme for subjects not at full risk of tumor development. The performance of the Poly-3 test depends on how closely it represents the correct specification of the time-at-risk weight in the data. Bailer and Portier further suggested that this test can be improved by using a general  $k$  reflecting the shape of the tumor onset distribution. In this article, we propose a method to estimate  $k$  by equating the empirical lifetime tumor incidence rate obtained from the data based on the fractional weighting scheme to a separately estimated cumulative lifetime tumor incidence rate. This Poly- $k$  test with the statistically estimated  $k$  appears to perform better than the Poly-3 test which is conducted without prior knowledge of the tumor onset distribution. Our simulation shows that the

proposed method improves the robustness to various tumor onset distributions in addition to the robustness to the effects of mortality achieved by the Poly-3 test. Large sample properties are shown via simulations to illustrate the consistency of the proposed method. The proposed methods are applied to analyze two real data sets. One is to find a dose-related linear trend on animal carcinogenicity, and the other is to test an effect of calorie restriction on experimental animals.

Key Words: Bioassay; Dose-response; Sacrifice; Trend test; Tumor incidence rate.

## 1 INTRODUCTION

Animal carcinogenicity studies often involve a problem of testing the statistical significance of a dose-response relationship among dose (treatment) groups. Many methods have been proposed for analyzing tumor incidence data from animal bioassays (see Ahn and Kodell [1]). Peto [2] proposed an approach that requires cause-of-death (COD) information for each animal to be determined by pathologists. This method assumes that pathologists clearly determine whether a tumor has affected an animal's risk of death and correctly classify each tumor as either incidental or fatal. The use of COD information in analysis of animal bioassays, however, has been controversial due to subjective decisions by pathologists. Misclassification of COD can produce biases for tests [3,4].

The Cochran-Armitage (CA) test [5,6] is introduced for detecting a linear trend across dose groups in the overall tumor rates. Although this test does not require COD information, it needs an assumption of equal risk for the tumor among the dose groups over the duration of the study. A problem for this test arises from the presence of treatment-induced mortality unrelated to the tumor of interest. The CA test is seen to be sensitive to these changes in treatment lethality [7-9]. To reduce this sensitivity, Bailer and Portier [7] proposed the Poly-3 test, which made an adjustment to the CA test by using a fractional weighting scheme for animals not at full risk of tumor development. They first mentioned the idea of the Poly- $k$  test without specifying how to obtain  $k$ . Since many tumors seem to appear at the rate of a third- to fifth-order polynomial in time (see Portier et al. [10]), they recommend the use of  $k = 3$ . The optimal value of  $k$  in the Poly- $k$  test relies on the shape of the tumor incidence function. As Bailer and Portier mentioned, if the shape of the tumor incidence function is expected to follow time to some power  $k$ , which is different from 3, then the Poly- $k$  test with  $k \neq 3$  should have superior operating characteristics to

the Poly-3 test. Therefore, the estimation of  $k$  in the Poly- $k$  test is a primary concern of this study.

In this paper, we propose a method for estimating  $k$  by equating a statistically estimated cumulative tumor incidence rate and an empirical estimate of tumor rate from the data based on the fractional weighting scheme. We use the term “Generalized Poly- $k$  test” to refer to a generalization of the Bailer and Portier’s Poly-3 test with an estimated  $k$ . According to our simulation study, the Generalized Poly- $k$  test provides robustness to a variety of tumor onset distributions as well as robustness to the effects of mortality benefited by the survival adjustment. Our method outperforms the conventional Poly-3 test when the true shape of the tumor incidence function is not close to the one with power 3.

Currently, our method is used for data with interim sacrifices. While the proposed method estimates  $k$  with a reasonably small bias for data with interim sacrifices, the estimation of tumor incidence rate for data with a single terminal sacrifice must be improved in order to achieve the desired accuracy. Since estimating the tumor incidence rate from single-sacrifice data requires additional assumptions [11,12], the estimation of  $k$  from single-sacrifice data is not as accurate as that for the data with interim sacrifices. Therefore, the estimation of  $k$  for data with a single terminal sacrifice is deferred to a future study.

The rest of the article is organized as follows. Section 2 reviews the Poly- $k$  trend test and a method for an estimation of tumor incidence rate used for estimating  $k$  in this paper. Section 3 proposes the method for estimating the tumor onset distribution. The proposed method is compared with the conventional Poly- $k$  test with fixed values of  $k$  at 1.5, 3 and 6 as well as the CA test ( $k = 0$ ) in Section 4. The consistency of the proposed method is also illustrated by simulations in large samples. Section 5 provides analyses of calorie restriction data and ED<sub>01</sub> data using the proposed method, and Section 6 presents our discussion.

## 2 SELECTED TUMOR INCIDENCE RATE ESTIMATION AND TREND TEST PROCEDURES

### 2.1 Estimation of tumor incidence rate

This section provides a review of some existing methods for the estimation of tumor incidence rate and for a test of dose-related trends for an animal carcinogenicity study. These procedures will be used to develop the proposed method for estimating the tumor onset distribution.

Kodell and Ahn [13] developed a method for estimating the tumor incidence rate for multiple-sacrifice experiments by modifying the nonparametric estimation method by Malani and Van Ryzin [14] in order to prevent a negative estimate on the tumor incidence rate. Kodell and Ahn's method is slightly modified and used to estimate the tumor onset distribution for the purpose of this study.

Consider an experiment with  $g$  treatment groups, where group  $i$  is exposed to a dose level  $d_i$  of a test substance, with  $d_1 = 0$  representing a control group. Divide the time scale into  $m$  intervals such that the  $j$ -th interval is  $(t_{j-1}, t_j]$ , where  $0 = t_0 < t_1 < \dots < t_m$ . A sacrifice is assumed to occur at the end of each time interval. Suppose  $N$  tumor-free animals are initially placed on the experiment, and  $N_i$  animals are assigned randomly to group  $i$ ,  $i = 1, \dots, g$ . While Kodell and Ahn estimate the tumor incidence rate for each dose group at each interval, we pool all the dose groups at each interval and estimate the tumor incidence rate in this study. At the  $j$ -th interval, for the pooled data, let  $a_{1j}$  and  $a_{2j}$  denote the numbers of pooled natural deaths with and without tumors, respectively, and  $s_{1j}$  and  $s_{2j}$  denote the numbers of pooled animals sacrificed with and without tumors, respectively.

Define two random variables  $X_T$  and  $X_D$  denoting, respectively, time to onset of tumor and time to natural death. The random variables  $X_T$  and  $X_D$  are discrete random variables, taking on only the values  $t_j$ ,  $j = 1, \dots, m$ . It is assumed that tumors precede natural deaths and deaths precede sacrifices. This assumption has been commonly made in practice (see Malani and Van Ryzin [14]; Kodell and Ahn [13]).

Let  $p_j^A = P(X_T \leq t_j | X_D > t_j)$  be the tumor prevalence function for live animals and  $p_j^D = P(X_T \leq t_j | X_D = t_j)$  the tumor prevalence function for dead animals. Define  $\lambda_j^D = P(X_D =$

$t_j|X_D \geq t_j$ ) to be the discrete hazard rate for  $X_D$ , and  $\lambda_j^S = P(X_S = j|X_S \geq j)$  to be the hazard function for time to sacrifice  $X_S$ . Then the log-likelihood function of  $p_j^A$ ,  $p_j^D$  and  $\lambda_j^D$  is given as

$$\begin{aligned} L \propto \sum_{j=1}^m & \left\{ a_{1j} \log p_j^D + a_{2j} \log(1 - p_j^D) + s_{1j} \log p_j^A + s_{2j} \log(1 - p_j^A) \right. \\ & + (a_{1j} + a_{2j}) \log \lambda_j^D + [B_j - (a_{1j} + a_{2j})] \log(1 - \lambda_j^D) \left. \right\} \\ & + \sum_{j=1}^{m-1} (s_{1j} + s_{2j}) \log \lambda_j^S + B_{j+1} \log(1 - \lambda_j^S), \end{aligned} \quad (1)$$

where  $B_j$  is the number of pooled animals alive at the beginning of the  $j$ -th interval, under the constraint that the tumor incidence rate for the  $j$ -th interval,

$$\begin{aligned} \lambda_j^T &= P(X_T = t_j | X_D \geq t_j, X_T \geq t_j) \\ &= 1 - \left\{ (1 - p_j^A) (1 - \lambda_j^D) + (1 - p_j^D) \lambda_j^D \right\} / (1 - p_{j-1}^A), \end{aligned} \quad (2)$$

is nonnegative. In order to estimate the tumor incidence rate, Kodell and Ahn [13,15] employed a numerical method for the constrained maximum likelihood estimation of  $\lambda_j^T$ .

## 2.2 Survival-adjusted trend test

Cochran [5] and Armitage [6] introduced a statistical test for detecting a linear trend across dose groups in lifetime tumor incidence rates. This test requires an assumption that all animals are at equal risk of developing a tumor over the duration of a study. The tumor data can be summarized as follows:

|                  | Dose level  |             |         |             | Total   |
|------------------|-------------|-------------|---------|-------------|---------|
|                  | $d_1$       | $d_2$       | $\dots$ | $d_g$       |         |
| # with tumors    | $y_1$       | $y_2$       | $\dots$ | $y_g$       | $y$     |
| # without tumors | $N_1 - y_1$ | $N_2 - y_2$ | $\dots$ | $N_g - y_g$ | $N - y$ |
| # subjects       | $N_1$       | $N_2$       | $\dots$ | $N_g$       | $N$     |

The Cochran-Armitage (CA) test utilizes the tumor data pooled over the entire duration of a study for each group.

Under the null hypothesis, the expected number of animals with tumors in the  $i$ -th group is  $E_i = yK_i$ , where  $K_i = N_i/N$ . Defining  $D_i$  as  $y_i - E_i$ , the test statistic for a possible monotonic trend with dose is based on  $X = \sum_{i=1}^g d_i D_i$  and the variance is estimated by

$$V = \frac{y(N-y)}{N(N-1)} \sum_{i=1}^g N_i (d_i - \bar{d})^2, \quad (3)$$

where  $\bar{d} = \sum_{i=1}^g N_i d_i / N$ . The CA test statistic is  $Z_{CA} = X / \sqrt{V}$ , where  $Z_{CA}$  is asymptotically distributed as a standard normal variate under the null hypothesis of equal tumor incidence rates among the groups. Despite the assumption for equal risk of getting the tumor during the study, some treatments shorten overall survival so that decreased risks of tumor onset may occur. The CA test is known to be sensitive to increases in treatment lethality and to fail to control the probability of a Type I error [7,9,16].

Bailer and Portier [7] proposed the Poly-3 test, which made an adjustment of the CA test by using a fractional weighting scheme for animals according to information on tumor presence/absence at the age of death. They define the number at risk for the  $i$ -th group as the sum of  $N_i$  weights

$$r_i = \sum_{h=1}^{N_i} w_{ih}, \quad (4)$$

where  $w_{ih}$  is the time-at-risk weight for the  $h$ -th animal in the  $i$ -th group. That is, the number of animals  $N_i$  in the CA test given in (3) is replaced with  $r_i$ , and  $r = \sum_i r_i$  is used in place of  $N$  in (3). The risk weight  $w_{ih}$  is defined as

$$w_{ih} = \begin{cases} 1 & \text{if the animal dies with the tumor} \\ \left(\frac{t_{ih}}{t_{\max}}\right)^k & \text{otherwise} \end{cases} \quad (5)$$

where  $t_{ih}$  ( $\leq t_{\max}$ ) is the actual death time of the animal, and  $t_{\max}$  is the time to termination of the experiment (e.g., terminal sacrifice time). This weighting scheme gives less weight to a tumor-free animal that dies at  $t_{ih} < t_{\max}$ . Bailer and Portier recommend that  $k = 3$  be used since previous work conducted by Portier et al. [10] in fitting Weibull hazards to tumor onset data showed that most tumors occur at the estimated rate of a third- to fifth-order polynomial in time.

Bieler and Williams [17] suggested a further modification of the CA test using the delta method and weighted least squares techniques to adjust the variance estimation of ratio statistics. With the notation introduced earlier, define  $\hat{p}'_i = y_i / r_i$  and  $\hat{p}' = y / \sum_i r_i$ ;  $a_i = r_i^2 / N_i$ ,  $v_{ih} = y_{ih} - \hat{p}' w_{ih}$ , and  $\bar{v}_i = \sum_h v_{ih} / N_i$ , where  $y_{ih}$  is the tumor indicator (1 for presence and 0 for absence) for the  $h$ -th animal in the  $i$ -th group. The test statistic of Bieler and Williams is

$$Z_{BW} = \frac{\sum_i a_i \hat{p}'_i d_i - (\sum_i a_i d_i)(\sum_i a_i \hat{p}'_i) / \sum_i a_i}{\sqrt{C \left\{ \sum_i a_i d_i^2 - (\sum_i a_i d_i)^2 / \sum_i a_i \right\}}}$$

where  $C = \sum_i \sum_h (v_{ih} - \bar{v}_i)^2 / (N - g)$ . This version of the Poly- $k$  test will be used throughout this paper.

### 3 ESTIMATION OF $K$

In order to find a suitable  $k$  for the Poly- $k$  test, we propose a method for estimating the tumor onset distribution. Among the animals from the entire experiment, let  $a_1$  and  $a_2$  denote the numbers of pooled natural deaths with and without tumors, respectively, and  $s_1$  and  $s_2$  denote the numbers of pooled animals sacrificed with and without tumors, respectively. On the basis of the fractional weighting scheme for subjects not at full risk of tumor development as described in Section 2.2, the empirical lifetime tumor incidence rate for the data may be obtained as

$$TI = \frac{a_1 + s_1}{\sum_{h=1}^N \{I_{A_h} + w_h(k)I_{B_h}\}}, \quad (6)$$

where  $w_h(k) = (t_h/t_{\max})^k$  is the time-at-risk weight for the  $h$ -th animal. Here,  $I_{A_h}$  and  $I_{B_h}$  are indicator functions such that  $A_h = \{\text{the } h\text{-th animal that died with a tumor}\}$  and  $B_h = \{\text{the } h\text{-th animal that died without a tumor}\}$ .

To estimate  $k$ , we equate (6) with the estimate of cumulative tumor incidence rate obtained by solving the MLE problem in (1) under the constraint in (2). In the present paper, therefore, the cumulative tumor incidence rate is estimated as

$$1 - \prod_{j=1}^m (1 - \hat{\lambda}_j^T), \quad (7)$$

where  $\hat{\lambda}_j^T$  is an estimate of the tumor incidence rate for the  $j$ -th interval from the pooled data across dose groups. Using the bisection method, the value of  $k$  equating (6) and (7) is obtained as the estimated  $k$  for the Generalized Poly- $k$  test. Note that the resulting test reduces to the CA test if  $k = 0$  is chosen.

### 4 SIMULATION STUDY

A Monte Carlo simulation study is conducted to evaluate the proposed Generalized Poly- $k$  test by comparing it with the Poly- $k$  test using a wide range of fixed  $k$  values.

A typical bioassay design with four dose groups of 50 animals each and an experimental duration of 2 years is used in this study according to standard designs of the National Toxicology Program (NTP). The design is simulated to have sacrifices at the end of NTP intervals [7], which are 52, 78, 92 and 104 weeks. In each group, six animals are randomly selected to be sacrificed at the end of each interval. All the remaining live animals are sacrificed at the end of the experiment. The dose levels used in the simulation are 0, 1, 2 and 4.

It is assumed that three independent random variables  $T_1$  (time to tumor onset),  $T_2$  (time from onset until death from the tumor), and  $T_3$  (time until death from a competing risk) completely determine the observed outcome for each animal. Distributions of time to onset and death are of the form used by Portier et al. [10]. The survival function of  $T_1$  is modeled as

$$S(t) = \exp \left\{ -\delta(t/t_{\max})^k \right\}, \quad (8)$$

where  $t_{\max}$  represents the duration of the study or the time for a terminal sacrifice. The value of  $k$  is set to be 1.5, 3 or 6 for the Weibull tumor onset distribution with a shape parameter of 1.5, 3 or 6, respectively. The best performance is expected from the Poly- $k$  test when the correct  $k$  is used for the data with the tumor onset distribution (8). The value of  $\delta$  is chosen such that the probability of tumor onset by the end of the experiment attains the desired rate. The tumor rates are chosen to be either 0.05, 0.15 or 0.30 for the control group. The tumor rates are set to be the same across dose groups for size evaluations. For power comparisons, the tumor rates at the highest dose group by 104 weeks are set to be 5, 3 and 2 times higher than the background tumor rates of 0.05, 0.15 and 0.30, respectively. The low and intermediate doses are chosen to be a quarter and a half of the highest dose, respectively.

The survival function for  $T_3$  is modeled as  $Q(t) = \exp \{-\phi(\gamma_1 t + \gamma_2 t^{\gamma_3})\}$ , where  $\phi \geq 1$ , and  $\gamma_1$ ,  $\gamma_2$  and  $\gamma_3$  are nonnegative. With  $\phi = 1$ ,  $\gamma_1 = 10^{-4}$  and  $\gamma_2 = 10^{-16}$ ,  $\gamma_3$  is calculated as  $\log[-\{\log Q_C(t_{\max}) + \gamma_1 t_{\max}\}/\gamma_2] / \log t_{\max}$  under the constraint that  $Q_C(t_{\max}) < \exp(-\gamma_1 t_{\max})$ , where  $Q_C(t_{\max})$  is the probability of survival with respect to competing risks in the control group at the end of the study. The value of  $\phi$  varies such that  $\phi = \log(\psi) / \log \{Q_C(t_{\max})\}$ , if the survival rate is  $\psi$ . The competing risks survival rate (CRSR) considered in this simulation is (0.7, 0.6, 0.5, 0.4), (0.5, 0.4, 0.3, 0.2), (0.6, 0.6, 0.6, 0.2) or (0.6, 0.6, 0.6, 0.6) for the control and three dose

groups. Several combinations of the CRSR are considered in order to represent most of the actual animal tumor experiments. In NTP feeding studies, the CRSR was about 50% for male F344 rats [18], but it was over 70% for mice and around 60% for female rats. The survival distribution for tumor-induced mortality,  $T_2$ , has the same form as the one for death from competing risks, and the values of  $\gamma_1$ ,  $\gamma_2$  and  $\gamma_3$  remain the same. The parameter  $\phi$  is selected from a database to reflect various tumor lethality. For each configuration in our simulation study, 10,000 simulated data sets are generated and a nominal significance level of  $\alpha = 0.05$  is used.

Tables 1 and 2 display the probability of a Type I error and power for the configurations considered in this simulation study. When CRSR is the same across dose groups, both the Generalized Poly- $k$  test and the Poly- $k$  tests with fixed  $k$  give Type I error rates that are reasonably close to the nominal significance level. The test results do not appear to be affected by the choice of  $k$  for these data. It is also illustrated graphically in Figure 1. This was expected because the intercurrent mortality patterns are equal across dose groups. Thus, we will focus on the configurations with different CRSR across dose groups.

The estimate of  $k$  tends to be higher for data with lower tumor rates. The value of  $k$  appears to be overestimated for data distributed with the Weibull tumor onset with a shape parameter of 1.5 ( $k = 1.5$ ), and underestimated for Weibull data with  $k = 6$ . The performance of our method is comparable with that of the Poly- $k$  test with the fixed correct value of  $k$ . Our method outperforms the Poly- $k$  test using a fixed, but incorrect, value of  $k$  in the sense of controlling the Type I error rate. Although the Poly-3 test yields reasonable size for data with  $k = 3$ , it tends to be anticonservative for data with  $k = 1.5$  and too conservative for data with  $k = 6$ . On the other hand, the proposed Generalize Poly- $k$  test performs well in all settings. Even for data with  $k = 3$ , the proposed method gave comparable results with the Poly-3 test which is suitable for this tumor onset distribution. Figure 1 also shows that the proposed test with estimated  $k$  performs as well as the Poly- $k$  test with true  $k$  in various intercurrent mortality patterns.

For all the configurations considered in these simulations with a wide range of tumor onsets and different intercurrent mortality patterns, our method appears to be quite robust. The Type I error rate obtained from our method ranges from 2.73 to 5.77, while that for the Poly- $k$  tests with fixed  $k$  at 1.5, 3, and 6 and the CA test ( $k = 0$ ) ranges within (1.15, 5.73), (1.80, 6.64), (3.73, 9.93) and (0.39, 5.56), respectively.

For data from the Weibull tumor onset distribution with  $k = 6$  and a tumor rate of 30%, the widely used Poly-3 test is quite conservative and could result in a significant loss of power (see Table 2). Although  $k$  is somewhat underestimated, the size and power of the Generalized Poly- $k$  test show a substantial improvement over the Poly-3 test. For data with  $k = 1.5$  and a tumor rate of 0.05, the Poly-3 test shows more inflated size compared to our method. On the other hand, the proposed test closely follows the Poly- $k$  test with a true value of  $k = 1.5$ , as shown in Figure 1.

Consistency for the proposed estimator is shown via a separate simulation study under a simulation design setting described earlier. Sample sizes of 50, 300, 500, and 1000 in each group and different numbers of sacrificed animals at the end of each time interval are considered. For each configuration, 1000 data sets are generated. Each data set consists of four groups with the same tumor rate. Since all the configurations in this simulation show similar pattern of improvement as the sample size increases, the estimated  $k$  and its standard error of selected configurations are reported. The average  $\hat{k}$  (and the standard error in parentheses) based on 1000 simulation data sets from different tumor onset distributions is reported in Table 3. Here, the data generated from tumor onset distribution with Weibull-1.5 have tumor rate 0.05 and CRSR (.5, .4, .3, .2); the data from Weibull-3 tumor onset distribution have tumor rate 0.05 and CRSR (.7, .6, .5, .4); and the data from Weibull-6 tumor onset distribution have tumor rate 0.15 and CRSR (.7, .6, .5, .4). It shows that our estimate of  $k$  would be close to the true value if sufficient information is provided through sacrifices. The bias is dramatically reduced when the sample size is increased up to 1000.

The accuracy of estimation is substantially improved if the competing risk is eliminated. Table 4 shows the simulation results for the data without competing risk. One thousand simulation data sets are generated for each configuration. For tumor onset distributions with Weibull-1.5 (tumor rate .05), Weibull-3.0 (tumor rate .05) and Weibull-6.0 (tumor rate .15), we obtain a nearly perfect estimation of  $k$  as the sample size increases. The accuracy of the estimation also depends on the number of sacrificed animals. Although enough information about the live animals is required from sacrifice, excessive sacrifice increases the bias. A design with an optimal number of sacrificed animals can be investigated in a future study.

## 5 EXAMPLES

### 5.1 $ED_{01}$ data: Trend test

An experiment was conducted at the National Center for Toxicological Research to study the effect of feeding 2-acetylaminofluorene (2-AAF) to female BALB/C mice ( $ED_{01}$  study [19]). In this paper, data from Room 142 is chosen for illustration. The tumors of interest are carcinomas of the bladder and hepatocellular adenomas or carcinomas of the liver. Table 5 contains frequencies of the groups of animals in Room 142 that were dosed continuously at concentrations of 0, 30, 35, 45, 60 and 75 ppm 2-AAF until the terminal sacrifice at 1001 days. These represent a subset of the whole experiment with control, 30, 35, 45, 60, 75, 100 and 150 ppm groups. Interim sacrifices were conducted at 273, 365, 420, 729 and 1001 days. The three low dose groups had less frequent interim sacrifices than did those in the 60 ppm and 75 ppm groups. Since some subjects without clear information on the status of a specific tumor were excluded from the data for each tumor site, the numbers of animals are not exactly matched for the two tumor sites.

Starting with the low dose groups, we gradually included higher dose groups in the trend tests because the liver tumor shows a significant dose-related trend for low dose groups. However, the bladder tumor does not show a significant trend in the low dose region. Table 6 provides the test results. When the control group is included in the trend test, a significant dose-related trend is detected for the liver tumor. On the other hand, it is not significant up to the inclusion of the 60 ppm group in the bladder tumor. Although the estimate of  $k$  is close to 3 for the liver tumor, it is much smaller for the bladder tumor. For trend tests with 35, 45, and 60 ppm groups for the liver tumor, the  $p$ -values are 0.058 for our method with  $\hat{k} = 2.0$ , 0.032 for Poly-3, and 0.392 for the Cochran-Armitage test. The Cochran-Armitage test gives a substantially different  $p$ -value from those of the Poly-3 test and the proposed method. The tests may cause a different decision between the proposed approach and the Poly-3 test if we use the popular nominal significant level of  $\alpha = 0.05$  in this case. Note that the estimated value of  $k$  is 2.0, which is closer to 1.5 than to 3, and the Poly-1.5 test (close to the estimated value) agrees with ours, giving  $p$ -value= 0.086. Thus, the implication is that, even though the true value of  $k$  is unknown, it appears to be 2.0 or less (our estimates are slightly biased upward for small  $k$ ). Hence, it seems that our test gives a more credible result than the Poly-3 test. Tests with the 30, 35 and 45 ppm groups for the liver tumor

show insignificant dose-related trends for both the proposed method and the Poly- $k$  tests, with all the different fixed values of  $k$ . For some choices of groups, the results between Poly-3 and ours are similar. Possible reasons might be because the data show a clear evidence one way or the other, or the CRSR are the same across dose groups.

## 5.2 *Calorie restriction data: Pairwise comparison*

Data from the Project on Calorie Restriction (PCR) conducted at the National Center for Toxicological Research are analyzed using the proposed method and compared with the existing tests. The PCR study was conducted to investigate the effects of calorie restriction on tumorigenicity and longevity in Fischer 344 rats [20]. For each sex, the study included both ad libitum (AL) and calorically restricted (CR) groups of animals. The study involved up to six scheduled sacrifices. The sacrifice times were at 368, 555, 754, 919, 1096 and 1292 days. All the animals were sacrificed or died before the 5-th interval in the male AL group and before the 6-th interval in the female AL group. Thus, we include the animals in the first four intervals for males and in the first five intervals for females. Except where indicated, each tumor type is analyzed for both males and females. In this paper, the tumors with a combined observed frequency of 5 or higher for a given sex are included. The tumor types included in this analysis are as follows:

- mononuclear cell leukemia (MCL)
- pituitary adenoma/carcinoma (PIT)
- islet adenoma/carcinoma (Islet)
- thyroid c-cell adenoma/carcinoma (C-Cell)
- pheochromocytoma benign/malignant and adrenal mixed malignant (PHEO)
- mammary fibroadenoma (FIB AD), females
- skin fibroma (FIBR), females
- interstitial cell adenoma (INTR), males
- clitoral gland adenoma/carcinoma (CLIT), females
- mesothelioma (MESO), males

The counts of sacrifices and deaths in each interval for each sex on different types of tumors are given in Figure 2. Tests on the effect of calorie restriction are carried out via the Generalized Poly- $k$

test and the Poly- $k$  tests with various values of fixed  $k$  for the above tumor types. Table 7 displays the  $p$ -values for the Generalized Poly- $k$  test and for the Poly- $k$  tests with fixed values of  $k$ . Among tumors considered in Figure 2, PIT, PHEO, FIB AD, INTR, and CLIT show a significant reduction of the tumor incidence rate in the CR group from one-sided statistical comparisons. On the other hand, MCL, Islet, C-Cell, FIBR, and MESO do not show a statistically significant difference in the comparison. In MCL and C-Cell tumors in female mice, it shows that the result from the Poly- $k$  test with fixed  $k = 6$  may be different from that from the other tests.

Although the Poly-3 and Generalized Poly- $k$  tests are in total agreement for the PCR data if a nominal significance level of 0.05 is used, their results differ slightly for some tumors if other levels are chosen. In such instances, it appears that the Generalized Poly- $k$  test gives the best results. For PHEO for females, for example,  $k$  is estimated to be 1.4. If the nominal significance level is 0.01, our test ( $p$ -value=0.016) agrees with the Poly-1.5 test ( $p$ -value=0.015) in not rejecting at the level of 0.01, while the Poly-3 test rejects ( $p$ -value=0.006) the null hypothesis. This is an example of how estimating  $k$  will help to reduce the anti-conservatism of the Poly-3 test when  $k$  is possibly smaller than 3.

## 6 DISCUSSION

The Poly-3 trend test eases the problem for the analysis of tumor incidence rates in the presence of treatment-induced mortality unrelated to the tumor of interest. Further, it does not require information on tumor lethality. However, the Poly-3 test is not expected to maintain the proper Type I error rate if the shape parameter of the tumor incidence function differs from 3. Although the Poly- $k$  test should have operating characteristics that are superior to the Poly-3 test when the optimal  $k$  is used, a statistical estimation of  $k$  has not been attempted in literature.

In this paper, we developed a statistical method for estimating  $k$  in order to enhance the robustness for the Poly- $k$  test to various shapes of the tumor incidence rate. Estimation of  $k$  was achieved by equating the empirical lifetime tumor incidence rate obtained from the data based on the fractional weighting scheme to a separately estimated cumulative lifetime tumor incidence rate. The performance and robustness were evaluated by Monte Carlo simulations considering various tumor onset distributions and various effects of mortality. The results indicate that the

proposed Generalized Poly- $k$  test performs well in all settings and is comparable to the Poly- $k$  test with correctly selected  $k$ . In addition, our simulation study on the data with increased sample sizes suggests that the proposed method is consistent provided that enough information is given through sacrifices in a large data set. The accuracy of the estimation also depends on the number of sacrificed animals. A design with an optimal number of sacrificed animals can be investigated in a future study.

As illustrated in analyzing data from the ED<sub>01</sub> study and the Project on Calorie Restriction, the Generalized Poly- $k$  test can result in more accurate assessment in finding the dose-related trend for a carcinogen (2-AAF) and the effect of calorie restriction on experimental animals.

Our results demonstrate that the estimation of  $k$  achieved the goal of providing robustness of the Poly- $k$  test to various choices of tumor incidence functions in addition to robustness to differential mortality as a treatment effect, which was already achieved. The Generalized Poly- $k$  test can be a good choice for analyzing data from animal carcinogenicity studies. The estimation of  $k$  for the data with a single terminal sacrifice is deferred to a future study. This remains an open research topic.

#### ACKNOWLEDGEMENTS

This work is supported in part by the Tobacco Settlement Funds approved by the Texas State Legislature (Hojin Moon and J. Jack Lee), NIH grants CA77289-05 (Hongshik Ahn) and CA91844 (J. Jack Lee).

#### REFERENCES

1. Ahn, H. and Kodell, R. L. Analysis of long-term carcinogenicity studies. In: *Design and analysis of animal studies in pharmaceutical development*, Chow, S. C. and Liu, J. P. (Eds.). Marcel Dekker, Inc.: New York, 1998; pp 259-289.
2. Peto, R. Guidelines on the the analysis of tumor rates and death rates in experimental animals. *British Journal of Cancer* 1974; **29**:101-105.

3. Lagakos, S. W. An evaluation of some two-sample tests used to analyze animal carcinogenicity experiments. *Utilitas Mathematica* 1982; **21B**:239-260.
4. Lagakos, S. W. and Louis, T. A. Use of tumour lethality to interpret tumorigenicity experiments lacking cause-of-death data. *Applied Statistics* 1988; **37**:169-179.
5. Cochran, W. G. Some methods for strengthening the common  $\chi^2$  tests. *Biometrics* 1954; **10**:417-451.
6. Armitage, P. Tests for linear trends in proportions and frequencies. *Biometrics* 1955; **11**:375-386.
7. Bailer, A. J. and Portier, C. J. Effects of treatment-induced mortality and tumor-induced mortality on tests for carcinogenicity in small samples. *Biometrics* 1988; **44**:417-431.
8. Kodell, R. L., Chen, J. J., and Moore, G. E. Comparing distributions of time to onset of disease in animal tumorigenicity experiments. *Communications in Statistics - Theory and Methods* 1994; **23**:959-980.
9. Moon, H, Ahn, H. and Kodell, R. L. Extension of Peto's test by attribution of tumor lethality in the absence of cause-of-death information. *Biometrical Journal* 2002; In press.
10. Portier, C., Hedges, J. and Hoel, D. G. Age-specific models of mortality and tumor onset for historical control animals in the National Toxicology Program's carcinogenicity experiments. *Cancer Research* 1986; **46**:4372-4378.
11. Lindsey, J. C. and Ryan, L. M. A comparison of continuous- and discrete-time three-state models for rodent tumorigenicity experiments. *Environmental Health Perspectives* 1994; **102** (Suppl. 1):9-17.
12. Kodell, R. L., Pearce, B. A., and Turturro, A. and Ahn, H. Age-adjusted trend test for the tumor incidence rate. *Drug Information Journal* 1997; **31**:471-487.
13. Kodell, R. L. and Ahn, H. An age-adjusted trend test for the tumor incidence rate. *Biometrics* 1997; **53**:1467-1474.

14. Malani, H. M. and Van Ryzin, J. Comparison of two treatments in animal carcinogenicity experiments. *Journal of the American Statistical Association* 1988; **83**:1171-1177.
15. Kodell, R. L. and Ahn, H. Nonparametric Trend Test for the Cumulative Tumor Incidence Rate. *Communications in Statistics - Theory and Methods* 1996; **25**:1677-1692.
16. Mancuso, J. Y., Ahn, H, Chen, J. J. and Mancuso, J. P. Age-adjusted exact trend tests in the event of rare occurrences. *Biometrics* 2002; In press.
17. Bieler G. S. and Williams R. L. Ratio estimates, the delta method, and quantal response tests for increased carcinogenicity. *Biometrics* 1993; **49**:793-801.
18. Haseman, J. K., Hailey, J. R. and Morris, R. W. Spontaneous neoplasm incidences in Fischer 344 rats and B6C3F<sub>1</sub> mice in two-year carcinogenicity studies: A National Toxicology Program Update. *Toxicologic Pathology* 1998; **26**:428-441.
19. Littlefield, N. A., Farmer, J. H., Gaylor, D. W. and Sheldon, W. G. Effects of Dose and Time in a Long-Term, Low-Dose Carcinogenic Study. *Journal of Environmental Pathology and Toxicology* 1980; **3**:17-34.
20. Thurman, J. D., Bucci, T. J., Hart, R. W. and Turturro, A. Survival, body weight, and spontaneous neoplasms in *ad libitum*-fed and food-restricted Fischer-344 rats. *Toxicologic Pathology* 1994; **22**:1-9.

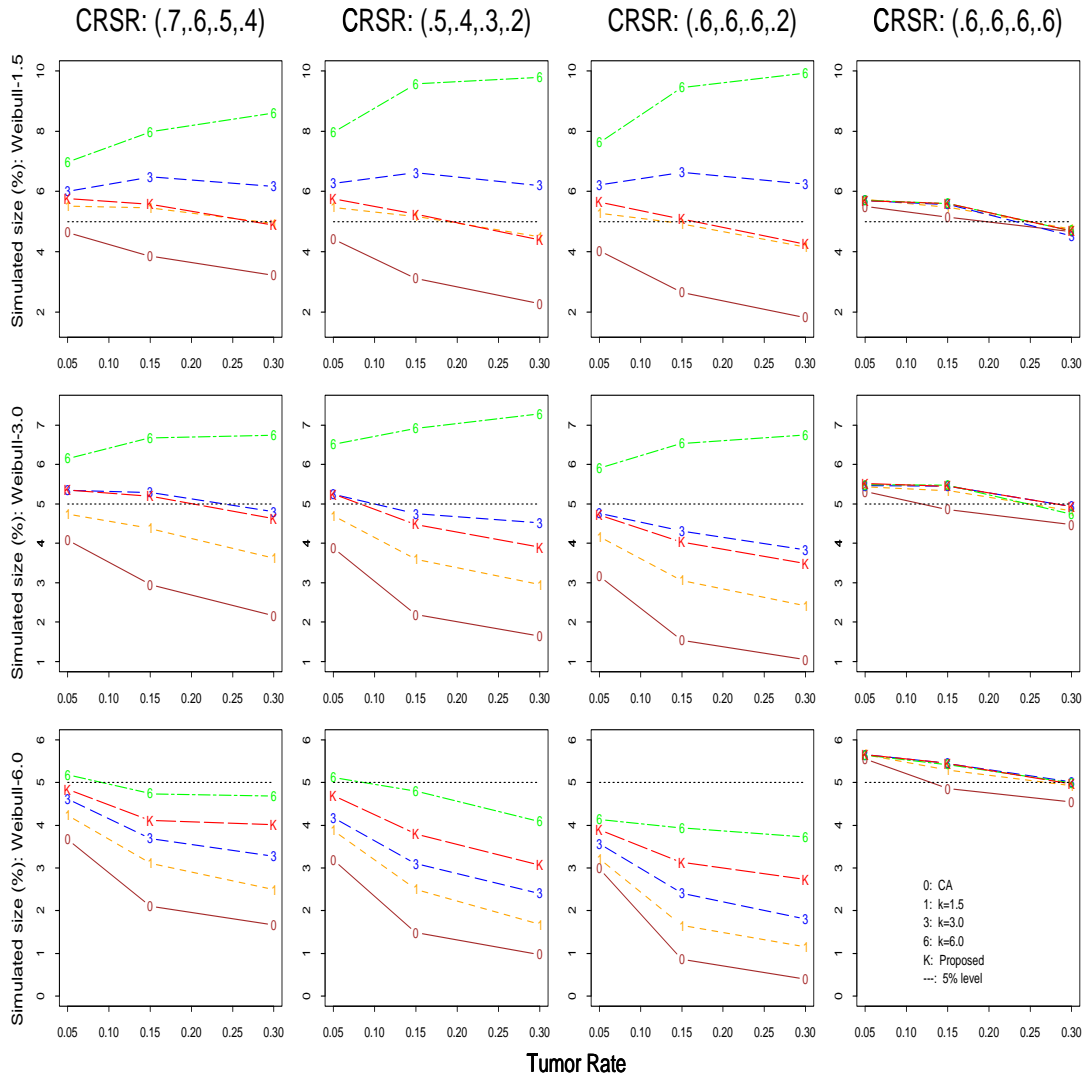


Figure 1: Simulated size of the proposed test and the Poly- $k$  test with various fixed values of  $k$  for various intercurrent mortality patterns and for several tumor onset distributions with various competing risks survival rates.

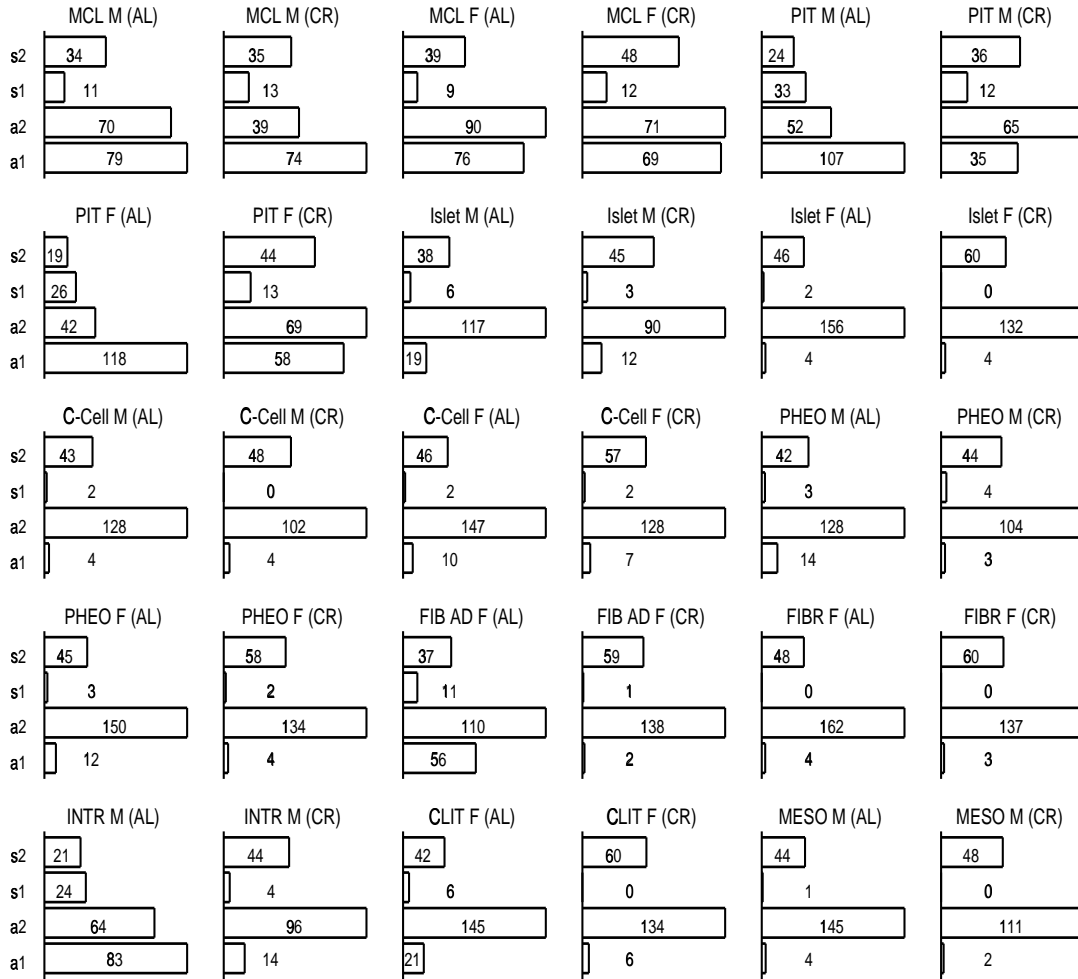


Figure 2: Counts for fitting the discrete model to data from the PCR experiment. Here, a1, a2, s1 and s2 represent #natural deaths with and without tumor and #sacrificed animals with and without tumor, respectively.

Table 1: Simulated size (%) and average  $\hat{k}$  corresponding to the nominal 5% significance level for the Generalized Poly- $k$  test, and the Poly- $k$  tests with several fixed  $k$  values. All tests are based on a dose scaling of 0, 1, 2, 4 with 50 animals per group and 10,000 trials for each configuration; approximate lethality is 35%.

| Tumor onset <sup>a</sup> | CRSR <sup>b</sup> | Tumor rate <sup>c</sup> | Poly- $k$ with fixed $k$    |           |         |         | Generalized Poly- $k$ |                                     |
|--------------------------|-------------------|-------------------------|-----------------------------|-----------|---------|---------|-----------------------|-------------------------------------|
|                          |                   |                         | CA <sup>d</sup> ( $k = 0$ ) | $k = 1.5$ | $k = 3$ | $k = 6$ | Size                  | Average $\hat{k}$ <sup>e</sup> (se) |
| 1.5                      | (.7 .6 .5 .4)     | .05                     | 4.64                        | 5.52      | 6.00    | 6.98    | 5.77                  | 2.44 (.019)                         |
|                          |                   | .15                     | 3.85                        | 5.45      | 6.48    | 7.98    | 5.57                  | 1.80 (.013)                         |
|                          |                   | .30                     | 3.21                        | 4.93      | 6.17    | 8.61    | 4.88                  | 1.61 (.010)                         |
|                          | (.5 .4 .3 .2)     | .05                     | 4.41                        | 5.47      | 6.27    | 7.97    | 5.76                  | 2.31 (.018)                         |
|                          |                   | .15                     | 3.11                        | 5.17      | 6.62    | 9.57    | 5.24                  | 1.66 (.012)                         |
|                          |                   | .30                     | 2.27                        | 4.49      | 6.20    | 9.79    | 4.39                  | 1.48 (.009)                         |
|                          | (.6 .6 .6 .2)     | .05                     | 4.03                        | 5.28      | 6.21    | 7.62    | 5.65                  | 2.40 (.019)                         |
|                          |                   | .15                     | 2.65                        | 4.93      | 6.64    | 9.45    | 5.08                  | 1.72 (.012)                         |
|                          |                   | .30                     | 1.81                        | 4.15      | 6.24    | 9.93    | 4.25                  | 1.55 (.009)                         |
|                          | (.6 .6 .6 .6)     | .05                     | 5.50                        | 5.73      | 5.70    | 5.71    | 5.70                  | 2.51 (.020)                         |
|                          |                   | .15                     | 5.14                        | 5.48      | 5.57    | 5.60    | 5.60                  | 1.77 (.013)                         |
|                          |                   | .30                     | 4.67                        | 4.75      | 4.53    | 4.70    | 4.70                  | 1.62 (.009)                         |
| 3.0                      | (.7 .6 .5 .4)     | .05                     | 4.09                        | 4.75      | 5.35    | 6.15    | 5.36                  | 3.49 (.021)                         |
|                          |                   | .15                     | 2.95                        | 4.38      | 5.29    | 6.68    | 5.20                  | 2.92 (.016)                         |
|                          |                   | .30                     | 2.16                        | 3.62      | 4.80    | 6.75    | 4.63                  | 2.74 (.012)                         |
|                          | (.5 .4 .3 .2)     | .05                     | 3.88                        | 4.70      | 5.25    | 6.51    | 5.25                  | 3.28 (.021)                         |
|                          |                   | .15                     | 2.19                        | 3.60      | 4.75    | 6.92    | 4.48                  | 2.65 (.015)                         |
|                          |                   | .30                     | 1.64                        | 2.95      | 4.52    | 7.29    | 3.90                  | 2.47 (.012)                         |
|                          | (.6 .6 .6 .2)     | .05                     | 3.17                        | 4.17      | 4.76    | 5.91    | 4.72                  | 3.43 (.021)                         |
|                          |                   | .15                     | 1.54                        | 3.06      | 4.31    | 6.54    | 4.03                  | 2.80 (.016)                         |
|                          |                   | .30                     | 1.05                        | 2.41      | 3.83    | 6.75    | 3.48                  | 2.64 (.012)                         |
|                          | (.6 .6 .6 .6)     | .05                     | 5.31                        | 5.43      | 5.46    | 5.49    | 5.52                  | 3.55 (.022)                         |
|                          |                   | .15                     | 4.86                        | 5.34      | 5.45    | 5.48    | 5.45                  | 2.91 (.016)                         |
|                          |                   | .30                     | 4.47                        | 4.82      | 4.94    | 4.74    | 4.93                  | 2.77 (.012)                         |
| 6.0                      | (.7 .6 .5 .4)     | .05                     | 3.67                        | 4.23      | 4.62    | 5.19    | 4.84                  | 4.87 (.021)                         |
|                          |                   | .15                     | 2.11                        | 3.11      | 3.69    | 4.74    | 4.11                  | 4.57 (.018)                         |
|                          |                   | .30                     | 1.66                        | 2.49      | 3.27    | 4.69    | 4.02                  | 4.59 (.015)                         |
|                          | (.5 .4 .3 .2)     | .05                     | 3.19                        | 3.88      | 4.17    | 5.12    | 4.70                  | 4.65 (.022)                         |
|                          |                   | .15                     | 1.48                        | 2.50      | 3.10    | 4.80    | 3.79                  | 4.20 (.018)                         |
|                          |                   | .30                     | 0.97                        | 1.68      | 2.40    | 4.09    | 3.07                  | 4.21 (.015)                         |
|                          | (.6 .6 .6 .2)     | .05                     | 2.99                        | 3.22      | 3.57    | 4.14    | 3.90                  | 4.85 (.021)                         |
|                          |                   | .15                     | 0.86                        | 1.65      | 2.41    | 3.94    | 3.13                  | 4.44 (.018)                         |
|                          |                   | .30                     | 0.39                        | 1.15      | 1.80    | 3.73    | 2.73                  | 4.45 (.015)                         |
|                          | (.6 .6 .6 .6)     | .05                     | 5.56                        | 5.66      | 5.65    | 5.65    | 5.66                  | 4.94 (.021)                         |
|                          |                   | .15                     | 4.86                        | 5.30      | 5.45    | 5.42    | 5.45                  | 4.60 (.018)                         |
|                          |                   | .30                     | 4.55                        | 4.93      | 5.02    | 4.99    | 4.97                  | 4.65 (.015)                         |

<sup>a</sup> $k$  in the Weibull tumor onset distribution function

<sup>b</sup>The four entries CRSR for control, low, intermediate and high dose groups, respectively

<sup>c</sup>Cumulative tumor onset probability (in control for power study) at 104 weeks in absence of competing risks

<sup>d</sup>Cochran-Armitage test

<sup>e</sup>Average value of the estimated  $k$  out of the 10,000 simulation data sets

Table 2: Simulated power (%) and average  $\hat{k}$  corresponding to the nominal 5% significance level for the Generalized Poly- $k$  test, and the tests with several fixed  $k$  values. All tests are based on a dose scaling of 0, 1, 2, 4 with 50 animals per group and 10,000 trials for each configuration; approximate lethality is 35%.

| Tumor onset <sup>a</sup> | CRSR <sup>b</sup> | Tumor rate <sup>c</sup> | Poly- $k$ with fixed $k$    |           |         |         | Generalized Poly- $k$ |                                     |
|--------------------------|-------------------|-------------------------|-----------------------------|-----------|---------|---------|-----------------------|-------------------------------------|
|                          |                   |                         | CA <sup>d</sup> ( $k = 0$ ) | $k = 1.5$ | $k = 3$ | $k = 6$ | Power                 | Average $\hat{k}$ <sup>e</sup> (se) |
| 1.5                      | (.7 .6 .5 .4)     | .05                     | 85.0                        | 88.6      | 89.7    | 91.1    | 88.5                  | 1.82 (.013)                         |
|                          |                   | .15                     | 90.3                        | 93.6      | 94.8    | 96.0    | 93.6                  | 1.56 (.010)                         |
|                          |                   | .30                     | 83.4                        | 89.3      | 91.3    | 93.4    | 89.3                  | 1.56 (.008)                         |
|                          | (.5 .4 .3 .2)     | .05                     | 80.8                        | 86.1      | 88.3    | 90.8    | 86.3                  | 1.73 (.013)                         |
|                          |                   | .15                     | 87.2                        | 92.2      | 94.1    | 95.9    | 91.8                  | 1.46 (.010)                         |
|                          |                   | .30                     | 77.5                        | 87.3      | 90.6    | 93.4    | 86.7                  | 1.44 (.008)                         |
|                          | (.6 .6 .6 .2)     | .05                     | 79.3                        | 85.7      | 88.2    | 90.6    | 85.9                  | 1.77 (.013)                         |
|                          |                   | .15                     | 85.3                        | 91.9      | 94.2    | 96.1    | 91.5                  | 1.52 (.010)                         |
|                          |                   | .30                     | 74.4                        | 87.0      | 90.8    | 94.1    | 86.5                  | 1.51 (.008)                         |
|                          | (.6 .6 .6 .6)     | .05                     | 88.6                        | 89.2      | 89.2    | 89.0    | 89.1                  | 1.87 (.014)                         |
|                          |                   | .15                     | 94.1                        | 94.4      | 94.3    | 93.9    | 94.4                  | 1.62 (.010)                         |
|                          |                   | .30                     | 89.9                        | 90.7      | 90.4    | 89.3    | 90.6                  | 1.61 (.008)                         |
| 3.0                      | (.7 .6 .5 .4)     | .05                     | 76.2                        | 81.2      | 82.8    | 85.4    | 82.6                  | 2.87 (.016)                         |
|                          |                   | .15                     | 82.4                        | 87.6      | 89.8    | 92.2    | 89.1                  | 2.65 (.013)                         |
|                          |                   | .30                     | 70.3                        | 79.8      | 83.8    | 87.9    | 83.1                  | 2.70 (.011)                         |
|                          | (.5 .4 .3 .2)     | .05                     | 68.5                        | 75.2      | 78.4    | 82.6    | 77.4                  | 2.66 (.016)                         |
|                          |                   | .15                     | 74.0                        | 82.6      | 86.2    | 90.2    | 84.4                  | 2.42 (.012)                         |
|                          |                   | .30                     | 59.3                        | 72.9      | 79.1    | 85.9    | 76.5                  | 2.45 (.011)                         |
|                          | (.6 .6 .6 .2)     | .05                     | 65.4                        | 73.9      | 77.7    | 81.9    | 76.7                  | 2.78 (.016)                         |
|                          |                   | .15                     | 70.2                        | 81.0      | 85.5    | 90.0    | 83.9                  | 2.57 (.012)                         |
|                          |                   | .30                     | 53.6                        | 70.8      | 78.3    | 85.7    | 76.3                  | 2.62 (.011)                         |
|                          | (.6 .6 .6 .6)     | .05                     | 83.9                        | 85.2      | 85.5    | 85.3    | 85.5                  | 2.98 (.017)                         |
|                          |                   | .15                     | 90.4                        | 91.4      | 91.5    | 91.2    | 91.5                  | 2.75 (.013)                         |
|                          |                   | .30                     | 84.6                        | 86.5      | 86.9    | 86.3    | 86.9                  | 2.80 (.011)                         |
| 6.0                      | (.7 .6 .5 .4)     | .05                     | 64.5                        | 70.5      | 72.8    | 76.1    | 74.4                  | 4.48 (.018)                         |
|                          |                   | .15                     | 67.1                        | 74.8      | 79.0    | 83.5    | 81.1                  | 4.39 (.014)                         |
|                          |                   | .30                     | 50.9                        | 62.3      | 68.7    | 76.1    | 73.0                  | 4.59 (.012)                         |
|                          | (.5 .4 .3 .2)     | .05                     | 51.8                        | 58.8      | 62.6    | 67.5    | 64.6                  | 4.16 (.019)                         |
|                          |                   | .15                     | 51.9                        | 62.7      | 68.9    | 76.1    | 71.2                  | 4.03 (.016)                         |
|                          |                   | .30                     | 34.1                        | 48.1      | 56.8    | 67.9    | 61.1                  | 4.16 (.014)                         |
|                          | (.6 .6 .6 .2)     | .05                     | 46.0                        | 55.7      | 60.0    | 65.7    | 62.7                  | 4.37 (.019)                         |
|                          |                   | .15                     | 44.9                        | 58.7      | 66.1    | 74.4    | 69.8                  | 4.30 (.016)                         |
|                          |                   | .30                     | 26.2                        | 43.2      | 53.7    | 66.2    | 60.4                  | 4.48 (.014)                         |
|                          | (.6 .6 .6 .6)     | .05                     | 77.7                        | 79.4      | 79.8    | 79.8    | 79.8                  | 4.62 (.018)                         |
|                          |                   | .15                     | 84.5                        | 85.8      | 86.3    | 86.6    | 86.4                  | 4.59 (.015)                         |
|                          |                   | .30                     | 76.8                        | 79.7      | 80.6    | 81.3    | 81.2                  | 4.77 (.013)                         |

<sup>a</sup>  $k$  in the Weibull tumor onset distribution function

<sup>b</sup> The four entries CRSR for control, low, intermediate and high dose groups, respectively

<sup>c</sup> Cumulative tumor onset probability (in control for power study) at 104 weeks in absence of competing risks

<sup>d</sup> Cochran-Armitage test

<sup>e</sup> Average value of the estimated  $k$  out of the 10,000 simulation data sets

Table 3: Average of  $\hat{k}$  (and se in parentheses) for the Generalized Poly- $k$  test from the simulated data with tumor rate .05, CRSR (.5, .4, .3, .2) for Weibull-1.5, tumor rate .05, CRSR (.7, .6, .5, .4) for Weibull-3.0, and tumor rate .15, CRSR (.7, .6, .5, .4) for Weibull-6.0. One thousand data sets are generated for each configuration. Each data set consists of four groups with the same tumor rate.

| Sample size | #Sac <sup>a</sup> | Tumor onset distribution |             |             |
|-------------|-------------------|--------------------------|-------------|-------------|
|             |                   | Weibull-1.5              | Weibull-3.0 | Weibull-6.0 |
| 50          | 6                 | 2.14 (.054)              | 3.29 (.066) | 3.93 (.051) |
|             | 10                | 2.10 (.055)              | 3.16 (.063) | 4.55 (.052) |
|             | 15                | 2.26 (.067)              | 3.32 (.075) | 4.82 (.063) |
| 300         | 36                | 1.31 (.026)              | 2.37 (.032) | 3.78 (.023) |
|             | 70                | 1.43 (.027)              | 2.69 (.036) | 4.78 (.030) |
|             | 90                | 1.62 (.034)              | 2.90 (.045) | 5.14 (.042) |
| 500         | 60                | 1.21 (.020)              | 2.31 (.025) | 3.77 (.018) |
|             | 120               | 1.32 (.020)              | 2.61 (.029) | 4.83 (.023) |
|             | 150               | 1.45 (.027)              | 2.73 (.038) | 5.17 (.034) |
| 1000        | 120               | 1.06 (.016)              | 2.19 (.020) | 3.76 (.014) |
|             | 250               | 1.22 (.016)              | 2.57 (.021) | 4.93 (.018) |
|             | 320               | 1.45 (.027)              | 2.76 (.037) | 5.36 (.035) |

<sup>a</sup>The number of animals in the interim sacrifice at each of the NTP interval

Table 4: Average of  $\hat{k}$  (and se in parentheses) for the Generalized Poly- $k$  test from the simulated data without competing risks. The tumor rates are chosen to be .05 (Weibull-1.5 and Weibull-3.0) and .15 (Weibull-6.0). One thousand data sets are generated for each configuration. Each data set consists of four groups with the same tumor rate.

| #/group <sup>a</sup> | Tumor onset distribution |                |             |                |             |                |
|----------------------|--------------------------|----------------|-------------|----------------|-------------|----------------|
|                      | Weibull-1.5              |                | Weibull-3.0 |                | Weibull-6.0 |                |
|                      | #Sac <sup>b</sup>        | $\hat{k}$ (se) | #Sac        | $\hat{k}$ (se) | #Sac        | $\hat{k}$ (se) |
| 300                  | 60                       | 1.73 (.027)    | 70          | 3.23 (.039)    | 90          | 5.99 (.035)    |
| 500                  | 110                      | 1.62 (.020)    | 140         | 3.11 (.034)    | 150         | 6.11 (.029)    |
| 1000                 | 200                      | 1.58 (.015)    | 270         | 3.07 (.024)    | 320         | 6.05 (.029)    |
| 1500                 | 340                      | 1.54 (.012)    | 460         | 3.01 (.027)    | 490         | 5.95 (.033)    |
| 2000                 | 610                      | 1.54 (.015)    | 630         | 3.00 (.026)    | 650         | 6.06 (.028)    |

<sup>a</sup> The number of animals in a group

<sup>b</sup> The number of animals in the interim sacrifice at each of the NTP interval

Table 5: Counts for fitting the discrete model to data from the ED<sub>01</sub> study.

| Dose   | Interval<br>(Day) | Bladder |         |         |         | Liver |       |       |       |
|--------|-------------------|---------|---------|---------|---------|-------|-------|-------|-------|
|        |                   | $a_1^a$ | $a_2^b$ | $s_1^c$ | $s_2^d$ | $a_1$ | $a_2$ | $s_1$ | $s_2$ |
| 0 ppm  | 0-273             | 1       | 4       | 0       | 24      | 0     | 5     | 0     | 24    |
|        | 274-365           | 0       | 8       | 0       | 24      | 0     | 8     | 0     | 24    |
|        | 366-420           | 0       | 4       | 0       | 23      | 0     | 4     | 0     | 23    |
|        | 421-729           | 1       | 118     | 0       | 137     | 3     | 121   | 1     | 136   |
|        | 730-1001          | 0       | 31      | 0       | 3       | 1     | 32    | 0     | 3     |
| 30 ppm | 0-273             | 1       | 7       | 0       | 0       | 0     | 10    | 0     | 0     |
|        | 274-365           | 1       | 9       | 0       | 0       | 1     | 9     | 0     | 0     |
|        | 366-420           | 0       | 6       | 0       | 0       | 0     | 6     | 0     | 0     |
|        | 421-729           | 6       | 317     | 0       | 281     | 21    | 314   | 24    | 257   |
|        | 730-1001          | 1       | 130     | 0       | 11      | 26    | 109   | 8     | 3     |
| 35 ppm | 0-273             | 0       | 9       | 0       | 0       | 0     | 12    | 0     | 0     |
|        | 274-365           | 1       | 2       | 0       | 0       | 1     | 3     | 0     | 0     |
|        | 366-420           | 0       | 4       | 0       | 0       | 0     | 3     | 0     | 0     |
|        | 421-729           | 0       | 219     | 1       | 201     | 14    | 217   | 17    | 185   |
|        | 730-1001          | 0       | 66      | 0       | 5       | 14    | 56    | 2     | 3     |
| 45 ppm | 0-273             | 0       | 7       | 0       | 0       | 0     | 6     | 0     | 0     |
|        | 274-365           | 0       | 2       | 0       | 0       | 0     | 3     | 0     | 0     |
|        | 366-420           | 0       | 1       | 0       | 0       | 0     | 1     | 0     | 0     |
|        | 421-729           | 3       | 137     | 0       | 156     | 11    | 140   | 21    | 135   |
|        | 730-1001          | 0       | 38      | 0       | 1       | 8     | 29    | 0     | 1     |
| 60 ppm | 0-273             | 0       | 4       | 0       | 48      | 0     | 4     | 0     | 48    |
|        | 274-365           | 0       | 2       | 0       | 46      | 0     | 2     | 0     | 46    |
|        | 366-420           | 0       | 4       | 0       | 38      | 1     | 3     | 0     | 40    |
|        | 421-729           | 1       | 116     | 2       | 159     | 7     | 118   | 30    | 131   |
|        | 730-1001          | 0       | 21      | 0       | 0       | 6     | 15    | 0     | 0     |
| 75 ppm | 0-273             | 0       | 4       | 0       | 24      | 0     | 6     | 0     | 24    |
|        | 274-365           | 0       | 5       | 0       | 23      | 0     | 5     | 0     | 23    |
|        | 366-420           | 1       | 2       | 0       | 22      | 0     | 3     | 0     | 22    |
|        | 421-729           | 4       | 95      | 1       | 115     | 22    | 81    | 26    | 90    |
|        | 730-1001          | 3       | 14      | 0       | 1       | 7     | 10    | 1     | 0     |

<sup>a</sup>Natural death with tumor

<sup>b</sup>Natural death without tumor

<sup>c</sup>Sacrificed with tumor

<sup>d</sup>Sacrificed without tumor

Table 6: Trend test results ( $p$ -values) for the Poly- $k$  test with fixed  $k$  and estimated  $k$  using the proposed method.

| Doses                   | Tumor   | Poly- $k$ with fixed $k$ |            |            |            | Proposed method |           |
|-------------------------|---------|--------------------------|------------|------------|------------|-----------------|-----------|
|                         |         | $k = 0$                  | $k = 1.5$  | $k = 3$    | $k = 6$    | $p$ -value      | $\hat{k}$ |
| (0, 30, 35)             | Bladder | .348                     | .161       | .436       | .472       | .495            | 0.6       |
|                         | Liver   | $\simeq 0$               | $\simeq 0$ | $\simeq 0$ | $\simeq 0$ | $\simeq 0$      | 2.8       |
| (0, 30, 35, 45)         | Bladder | .340                     | .424       | .456       | .465       | .382            | 0.6       |
|                         | Liver   | $\simeq 0$               | $\simeq 0$ | $\simeq 0$ | $\simeq 0$ | $\simeq 0$      | 2.6       |
| (0, 30, 35, 45, 60)     | Bladder | .448                     | .429       | .418       | .384       | .444            | 0.2       |
|                         | Liver   | $\simeq 0$               | $\simeq 0$ | $\simeq 0$ | $\simeq 0$ | $\simeq 0$      | 2.2       |
| (0, 30, 35, 45, 60, 75) | Bladder | .014                     | .008       | .006       | .003       | .008            | 1.6       |
|                         | Liver   | $\simeq 0$               | $\simeq 0$ | $\simeq 0$ | $\simeq 0$ | $\simeq 0$      | 2.2       |
| (30, 35, 45)            | Bladder | .705                     | .696       | .685       | .649       | .701            | 0.6       |
|                         | Liver   | .289                     | .269       | .246       | .186       | .256            | 2.4       |
| (35, 45, 60)            | Bladder | .304                     | .211       | .176       | .139       | .286            | 0.2       |
|                         | Liver   | .392                     | .086       | .032       | .009       | .058            | 2.0       |

Table 7: Test results ( $p$ -values) for the Poly- $k$  test with fixed  $k$  and estimated  $k$  using the proposed method.

| Tumor  | Sex | Poly- $k$ with fixed $k$ |            |            |            | Proposed method |           |
|--------|-----|--------------------------|------------|------------|------------|-----------------|-----------|
|        |     | $k = 0$                  | $k = 1.5$  | $k = 3$    | $k = 6$    | $p$ -value      | $\hat{k}$ |
| MCL    | M   | .924                     | .902       | .812       | .526       | .856            | 2.4       |
|        | F   | .436                     | .401       | .165       | .010       | .220            | 2.6       |
| PIT    | M   | $\simeq 0$               | $\simeq 0$ | $\simeq 0$ | $\simeq 0$ | $\simeq 0$      | 1.0       |
|        | F   | $\simeq 0$               | $\simeq 0$ | $\simeq 0$ | $\simeq 0$ | $\simeq 0$      | 1.0       |
| Islet  | M   | .141                     | .097       | .057       | .017       | .083            | 2.0       |
|        | F   | .293                     | .227       | .157       | .068       | .222            | 1.6       |
| C-Cell | M   | .337                     | .294       | .242       | .156       | .216            | 3.8       |
|        | F   | .294                     | .203       | .114       | .028       | .171            | 2.0       |
| PHEO   | M   | .050                     | .031       | .017       | .005       | .024            | 2.2       |
|        | F   | .030                     | .015       | .006       | .001       | .016            | 1.4       |
| FIB AD | F   | $\simeq 0$               | $\simeq 0$ | $\simeq 0$ | $\simeq 0$ | $\simeq 0$      | 0.6       |
| FIBR   | F   | .386                     | .326       | .255       | .147       | .255            | 3.0       |
| INTR   | M   | $\simeq 0$               | $\simeq 0$ | $\simeq 0$ | $\simeq 0$ | $\simeq 0$      | 0.8       |
| CLIT   | F   | $\simeq 0$               | $\simeq 0$ | $\simeq 0$ | $\simeq 0$ | $\simeq 0$      | 0.6       |
| MESO   | M   | .184                     | .160       | .133       | .090       | .143            | 2.4       |