

Extension of Peto's Test by Attribution of Tumor Lethality in the Absence of Cause-of-Death Information

Hojin Moon^{1*}, Hongshik Ahn² and Ralph L. Kodell³

¹Biostatistics, UT MD Anderson Cancer Center, 1515 Holcombe Blvd., Houston, TX 77030-4009

²Dept. of Appl. Math and Statistics, State University of New York, Stony Brook, NY 11794-3600

³Division of Biometry and Risk Assessment, NCTR, USFDA, Jefferson, AR 72079

Abstract: A new statistical testing approach is developed for rodent tumorigenicity assays that have a single terminal sacrifice or occasionally interim sacrifices but not cause-of-death data. For experiments that lack cause-of-death data, statistically imputed numbers of fatal tumors and incidental tumors are used to modify Peto's cause-of-death test which is usually implemented using pathologist-assigned cause-of-death information. The numbers of fatal tumors are estimated using a constrained nonparametric maximum likelihood estimation method. A new Newton-based approach under inequality constraints is proposed for finding the global maximum likelihood estimates. In this study, the proposed method is concentrated on data with a single sacrifice experiment without implementing further assumptions. The new testing approach may be more reliable than Peto's test because of the potential for a misclassification of cause-of-death by pathologists. A Monte Carlo simulation study for the proposed test is conducted to assess size and power of the test. Asymptotic normality for the statistic of the proposed test is also investigated. The proposed testing approach is illustrated using a real data set.

Keywords: bioassay; competing risk; fatal tumor; incidental tumor; sacrifice.

*Corresponding author: hmoon@mail.mdanderson.org

1 Introduction

In survival/sacrifice experiments, information concerning onset of specific diseases is confounded with information concerning the effect of the presence or absence of all diseases on mortality. In order to avoid biases due to differences in intercurrent mortality, Peto (1974) and Peto et al. (1980) proposed a method for analyzing tumor data in which tumors are observed in both the fatal and incidental contexts. Tumors that do not alter an animal's risk of death and are observed only as the result of a death from an unrelated cause are classified as incidental, whereas tumors that affect mortality by either directly causing death or indirectly increasing the risk of death from other causes are classified as fatal. This segregation allows appropriate numbers at risk to be used in the calculation of age-adjusted tests for tumorigenic effects of test agents. The analysis of data on occult tumors using cause of death (COD) is performed separately for nonlethal (incidental) tumors and lethal (fatal) tumors. This widely used statistical test (Peto et al., 1980) is recommended by the International Agency for Research on Cancer (IARC).

Many analyses require COD information (Peto, 1974; Kodell and Nelson, 1980; Peto et al., 1980; Kodell, Shaw and Johnson, 1982; Turnbull and Mitchell, 1984; Kodell and Chen, 1987; Archer and Ryan, 1989). These methods are popular because they do not require large numbers of animals to be sacrificed at multiple timepoints to observe the prevalence of occult tumors. These statistical methods assume that pathologists can determine if a tumor affected an animal's risk of death. In practice, however, pathologists often claim that accurate determination of the COD is not feasible, and classification errors can produce biases (Lagakos, 1982; Racine-Poon and Hoel, 1984; Lagakos and Louis, 1988).

An analysis of the COD assignment in selected rodent carcinogenicity studies at the National Center for Toxicological Research (NCTR) indicated that pathologists were inclined to assign a single lesion as the probable COD for most dead/moribund animals (Kodell et al., 1995). Unfortunately, the analysis could not assess the degree of accuracy of COD assignment in those studies.

Due to the COD controversy, researchers have proposed various alternative statistical tests that do not rely on COD information. Usually such tests require either interim-sacrifice data (McKnight and Crowley, 1984; Dewanji and Kalbfleisch, 1986; Malani and Van Ryzin, 1988; Gómez and Van Ryzin, 1992) or additional assumptions (Bailer and Portier, 1988; Bieler and Williams, 1993).

Among these, Bailer and Portier (1988) proposed the Poly-3 trend test, which made an adjustment of the Cochran-Armitage (CA) trend test (Cochran, 1954; Armitage, 1955) for detecting a linear trend across dose groups in the overall proportions of animals with tumor. Bieler and Williams (1993) made a further adjustment to the CA test. Currently, most of the animal tumorigenicity studies are designed with a single terminal sacrifice. Lindsey and Ryan (1994) and Kodell et al. (1997) proposed statistical methods for analyzing tumor incidence rates based on data with a single terminal sacrifice. These methods require further assumptions.

Peto's test requires data with COD assigned by pathologists. However, as mentioned earlier, COD information is not always available and is subject to error. Thus, our primary goal is to develop a new method of making Peto's test available for data without COD information. This can be done by replacing the COD assigned by pathologists with a statistically imputed COD. The numbers of fatal and incidental tumors imputed by a modification of the constrained nonparametric maximum likelihood estimation (NPMLE) method by Ahn et al. (2000) are used in the proposed test. Instead of using the lethality function introduced in Ahn et al., all possible integer pairs of the number of fatal and incidental tumors given the number of deaths with tumor from the data are examined in the maximum likelihood estimation. The value with the largest maximum likelihood is chosen as the estimated number of fatal tumors. The asymptotic normality of the proposed test statistic is examined via normal probability plots. We also conduct a comparison of our test statistic with Peto's test statistic which is proven to be asymptotically normal.

In this paper, a Newton-based approach with a quadratic convergence rate is proposed to find the NPMLE. The proposed algorithm is easy to implement and substantially faster than the Complex Method used in Ahn et al. (see also Moon et al., 1999). Ahn et al. (2000) developed the estimation methods for multiple-sacrifice data. However, this paper is concentrated on applying the proposed methods for data with single terminal sacrifice without additional assumptions. The new testing procedure is implemented in C and Fortran, and programs are available from the authors.

The present paper extends and improves upon the methods in Ahn et al. (2000) by (1) modifying the likelihood function to find a better estimate of the tumor lethality, and (2) extending the availability of the method to single terminal sacrifice data as well as multiple sacrifice data. Furthermore, this paper proposes the following new methods in estimation and testing: (1) A new Newton-based approach under inequality constraints is used for finding the global constrained

maximum likelihood estimates. (2) Peto's test is modified for data lacking COD information.

The proposed procedures are used to test the effect of feeding 2-AAF to female mice from the ED₀₁ study conducted at the NCTR, using a subset of the data which includes only a single terminal sacrifice. The main interest in this project is to investigate the dose-related trend of the liver tumor incidence.

2 Attribution of Tumor Lethality

We modify the Peto test for data without cause-of-death information. Our modified Peto test uses statistically attributed cause of death rather than the information given by pathologists.

Consider an experiment with g treatment groups, a control and $g - 1$ dose groups. Suppose that N_i animals in the i th group are followed over time for the development of irreversible and occult tumors. We assume that all animals come from the same population and are born without tumor on day zero of the experiment. Let $0 = t_0 < t_1 < \dots < t_J$ be $J + 1$ appropriately spaced time points, where t_J denotes the terminal sacrifice time. Divide the time scale into J intervals such that the j th interval is given by $I_j = (t_{j-1}, t_j]$, $j = 1, \dots, J$.

The basic terminology and notation of Ahn et al. (2000) are used here with some modification. Define $S(t)$ as the survival function for the distribution of time to tumor onset, i.e., $S(t) = \Pr(T_1 > t)$, where T_1 is the random variable representing time to onset of the tumor of interest. Define $P(t)$ as the survival function with respect to death caused by the tumor of interest, i.e., $P(t) = \Pr(T_D > t)$, where T_D represents the overall time to death from the tumor of interest. Define $Q(t)$ as the survival function with respect to death from competing risks, i.e., $Q(t) = \Pr(X_C > t)$, where X_C is the random variable representing time to death from a cause other than the tumor of interest. Figure 1 illustrates a tumor-death model introduced by Kodell and Nelson (1980). We assume that probabilities involving T_1 , T_D and X_C can only be evaluated at the ends of intervals, because we have only interval data (interval support). We also assume that T_1 and T_D occur before X_C in an interval. These assumptions are commonly made in practice (see Malani and Van Ryzin, 1988; Kodell and Ahn, 1997). The contributions to the likelihood function in Ahn et al. are slightly modified and formulated in terms of $S(t)$, $P(t)$ and $Q(t)$ as if complete information on COD is available for the experimental animals.

To develop the likelihood contributions, we begin by assuming that complete data on cause of death are available. For a given experimental group, let d_j denote the number of animals observed with fatal tumors in the j th interval. Let a_{ij} denote the number of animals in the j th interval observed to die with incidental tumors and b_{1j} the number observed without tumors. Let a_{2j} and b_{2j} denote similar quantities for sacrificed animals.

The probability that each of d_j animals dies due to the tumor of interest in the j th interval is $\Pr[t_{j-1} < T_D \leq t_j, X_C > T_D] = \Pr[t_{j-1} < T_D \leq t_j] \Pr[X_C > T_D \mid t_{j-1} < T_D \leq t_j]$. Given that $t_{j-1} < T_D \leq t_j$, and T_D must occur before X_C within an interval, X_C could have also occurred in the j th interval after T_D . Thus, for $X_C > T_D$, all we can be sure of is that $X_C > t_{j-1}$, i.e., $\Pr[X_C > T_D \mid t_{j-1} < T_D \leq t_j] = \Pr[X_C > t_{j-1}] = Q(t_{j-1})$. Therefore, the likelihood contribution is given as $Q(t_{j-1})\{P(t_{j-1}) - P(t_j)\}$.

Analogously, the probability that each of the a_{1j} animals dies in the j th interval from unrelated causes but has the tumor of interest is $\Pr[(T_D > X_C, T_1 \leq X_C) \text{ and } (t_{j-1} < X_C \leq t_j)] = \Pr[t_{j-1} < X_C \leq t_j] \Pr[(T_D > X_C, T_1 \leq X_C) \mid t_{j-1} < X_C \leq t_j]$. Now, if T_1 and T_D must occur before X_C in an interval, then given $t_{j-1} < X_C \leq t_j$, for T_D to be greater than X_C , T_D cannot occur in the j th interval. Thus, we can be sure that $T_D > t_j$. However, T_1 can occur before X_C in the j th interval and still satisfy $T_1 \leq X_C$. Hence, $\Pr[(T_D > X_C, T_1 \leq X_C) \mid t_{j-1} < X_C \leq t_j] = \Pr[T_D > t_j, T_1 \leq t_j] = P(t_j) - S(t_j)$. Therefore, the likelihood contribution is given as $\{Q(t_{j-1}) - Q(t_j)\} \{P(t_j) - S(t_j)\}$.

The probability that each of the b_{1j} animals dies without the tumor of interest in the j th interval is $\Pr[T_1 > X_C, t_{j-1} < X_C \leq t_j] = \Pr[t_{j-1} < X_C \leq t_j] \Pr[T_1 > X_C \mid t_{j-1} < X_C \leq t_j]$. If T_1 must occur before X_C in an interval, then for T_1 to be greater than X_C given $t_{j-1} < X_C \leq t_j$, T_1 cannot occur in the j th interval. Then $T_1 > X_C$ implies $T_1 > t_j$. Thus, $\Pr[T_1 > X_C \mid t_{j-1} < X_C \leq t_j] = \Pr[T_1 > t_j] = S(t_j)$. Therefore, the likelihood contribution can be derived as $S(t_j)\{Q(t_{j-1}) - Q(t_j)\}$.

For the sacrifice data, the probability that each of the a_{2j} animals sacrificed at t_j has the tumor of interest is $\Pr[T_1 \leq t_j, T_D > t_j] \Pr[X_C > t_j]$, and $Q(t_j)\{P(t_j) - S(t_j)\}$ is contributed to the likelihood function. The probability that each of the b_{2j} animals sacrificed does not have the tumor at t_j is $\Pr[T_1 > t_j] \Pr[X_C > t_j]$, and $S(t_j)Q(t_j)$ is contributed to the likelihood. For experiments with a single terminal sacrifice, a_{2j} and b_{2j} are zero except for the last interval. This formulation of the likelihood contributions is the same as that in Kodell and Chen (1987).

Ahn et al. introduced the lethality function $\lambda_j = \Pr(\text{tumor causes death} \mid \text{death with tumor})$. From the formula, $\lambda_j = 0$ as expected for nonlethal tumors since $p_j = 1$. When the tumor is instantly lethal, $\lambda_j = 1/(2 - q_j)$ is obtained since $\pi_j = 1$, and consequently, λ_j ranges from 0.5 to 1 depending on the competing risks survival rate. Although the estimation method in this paper does not require λ_j , if formulated in terms of the present likelihood contributions, it gives value 0 for incidental tumors and 1 for fatal tumors.

For fixed d_j and a_{1j} , a new likelihood function for an individual dose group in experiments lacking COD data is given as

$$L \propto \prod_{j=1}^J \{Q(t_{j-1})[P(t_{j-1}) - P(t_j)]\}^{d_j} \{[Q(t_{j-1}) - Q(t_j)][P(t_j) - S(t_j)]\}^{a_{1j}} \\ \times \{S(t_j)[Q(t_{j-1}) - Q(t_j)]\}^{b_{1j}} \{Q(t_j)[P(t_j) - S(t_j)]\}^{a_{2j}} \{S(t_j)Q(t_j)\}^{b_{2j}}.$$

For all j , let $\pi_j = S(t_j)/P(t_j)$, $p_j = P(t_j)/P(t_{j-1})$ and $q_j = Q(t_j)/Q(t_{j-1})$. Then the survival functions can be expressed as

$$P(t_j) = \prod_{k=1}^j p_k, \quad Q(t_j) = \prod_{k=1}^j q_k \quad \text{and} \quad S(t_j) = \pi_j P(t_j), \quad j = 1, \dots, J. \quad (1)$$

After the reparameterization, the log-likelihood function is, apart from a constant, derived as

$$l = \sum_{j=1}^J \left\{ (N_{j-1} - N_j) \sum_{k=1}^{j-1} \log(p_k q_k) + (a_{1j} + b_{1j} + a_{2j} + b_{2j}) \log p_j \right. \\ \left. + d_j \log(1 - p_j) + (a_{1j} + b_{1j}) \log(1 - q_j) \right. \\ \left. + (a_{2j} + b_{2j}) \log q_j + (b_{1j} + b_{2j}) \log \pi_j + (a_{1j} + a_{2j}) \log(1 - \pi_j) \right\}, \quad (2)$$

where N_j is the number of live animals at t_j .

In this problem, d_j and a_{1j} need to be imputed since they are not available due to the lack of COD information, although $d_j + a_{1j}$ can be observed from the data. The imputation procedure is as follows: All possible integer values of (a_{1j}, d_j) given $d_j + a_{1j} =$ (the number of animals that died in interval j with the tumor of interest) are considered in the estimation. Function l is maximized for each pair of (d_j, a_{1j}) . Thus, $d_j + a_{1j} + 1$ maximizations are considered. The integer values of (d_j, a_{1j}) with the largest maximum likelihood and the corresponding MLE's are chosen as the desired estimates. This approach can be viewed as an EM algorithm by assuming that the number of deaths with fatal tumor, d_j , and the number of deaths with incidental tumor, a_{1j} , are predictable

from the information on the number of animals dying with tumor, $d_j + a_{1j}$. Therefore, $(\hat{d}_j, \hat{a}_{1j})$ should be one of all possible integer pairs of (d_j, a_{1j}) given fixed $d_j + a_{1j}$.

Due to the large number of parameters to be estimated for each experimental group, a sufficient number of animals per group is required. The design used by the NTP currently incorporates 50 animals per group in a standard 2-year study (Bieler and Williams, 1993).

The monotonicity of $S(t)$, $P(t)$ and $Q(t)$, and the condition $S(t) \leq P(t)$ provide the following inequality conditions:

$$0 \leq \pi_j \leq 1, 0 \leq p_j \leq 1, 0 \leq q_j \leq 1, \text{ and } \pi_j p_j \leq \pi_{j-1} \text{ for } j = 1, \dots, J. \quad (3)$$

The MLE's of π_j , p_j and q_j can be obtained by maximizing (2) subject to (3).

Among all dose groups, define the tumor-death-ratio as the ratio of the number of natural deaths with tumor and the total number of natural deaths, and the tumor-sacrifice-ratio as the ratio of the number of animals having the tumor of interest among sacrificed animals and the total number of sacrificed animals. One way to determine the degree of lethality is comparing tumor-death-ratio and tumor-sacrifice-ratio. A high tumor-death-ratio compared to the tumor-sacrifice-ratio implies a highly lethal tumor. A ratio $\gamma_T = \text{tumor-death-ratio}/\text{tumor-sacrifice-ratio}$ is calculated to impose an upper bound on d_j as $(d_j + a_{1j})(\gamma_T/\gamma_{\max})$ for $\gamma_T < \gamma_{\max}$, and $d_j + a_{1j}$ otherwise, where γ_{\max} is a given upper bound of γ_T . In this study, γ_{\max} is set to be 2. According to the simulation, our method estimates d_j quite well for data with moderate and high lethality, but d_j tends to be overestimated for low lethal tumors. Therefore, this adjustment reduces the bias of the estimation for data with low lethal tumors. For highly lethal tumors, the bias is not expected to increase substantially because γ_T is almost always higher than 2. Even if the ratio is less than 2, d_j can go close to $a_1 + d$ if the ratio is not much lower than 2. An upper bound of 2 improves estimation for low lethal tumors without damaging accuracy of estimation for data with higher lethality according to our simulation study. An upper bound of 1 is not enough to reduce the bias since we expect $\gamma_T \simeq 1$ for incidental tumors.

When the Newton method is used for the current optimization problem, modifications are necessary to satisfy the bounds on the parameters and the complicated nonlinear constraints. The Newton-based method given in Ahn et al. (2002) is employed for finding the maximum of a new

log-likelihood function given in (2) under the constraints of (3). This method is easy to implement and faster than the direct-search algorithm used in Ahn et al. (2000) by the order of magnitude 1000. A quadratic convergence rate of the Newton Method is possible to obtain with a properly chosen initial guess which resides in the domain of contraction. The survival functions can be estimated by substituting the maximum likelihood estimates into (1).

3 Modified Peto's Test via Attribution of Tumor Lethality

In this section, a modified Peto test for data with unknown COD is proposed. As in the Peto test, our test assumes no treatment effect among the g groups under the null hypothesis. A dose-related trend in the tumor incidence rate is tested in the proposed method. By using \hat{d}_j, \hat{a}_{1j} ($j = 1, \dots, J$) obtained in Section 2, Peto's test can be modified for data lacking COD information. A detailed description of the Peto test can be found in Peto (1974). See also Ahn and Kodell (1998). In this study, the NTP (National Toxicology Program) intervals (Bailer and Portier, 1988) are used. Instead of using a fixed length, the NTP intervals reflect the number of observed animals by reducing the length of the interval as time goes by. See Section 4.1 for further details on these intervals.

First, consider Table 1 which summarizes tumor prevalence data for interval I_j . Let $n_{ij} = \hat{a}_{1ij} + a_{2ij} + b_{1ij} + b_{2ij}$ be the number of animals in group i dying during interval I_j from causes unrelated to the presence of the tumor of interest, and $y_{ij} = \hat{a}_{1ij} + a_{2ij}$ be the number of these animals in which the tumor was observed in the incidental context, for $i = 1, \dots, g$. All tumors found in sacrificed animals are classified as incidental.

The expected number of tumors in the i th group for the j th interval is $E_{ij} = y_{.j}K_{ij}$, where $K_{ij} = n_{ij}/n_{.j}$. Thus, the observed and expected numbers of tumors in the i th group over the entire experiment are $O_i = \sum_{j=1}^m y_{ij}$ and $E_i = \sum_{j=1}^m E_{ij}$, respectively, for $i = 1, \dots, g$. Define $D_i = O_i - E_i$ and V_{ri} as

$$D_i = \sum_{j=1}^m (y_{ij} - E_{ij}) \quad \text{and} \quad V_{ri} = \sum_{j=1}^m \kappa_j K_{rj} (\delta_{ri} - K_{ij}),$$

where $\kappa_j = y_{.j}(n_{.j} - y_{.j})/(n_{.j} - 1)$ and δ_{ri} is defined as 1 if $r = i$ and 0 otherwise. Let $\mathbf{D}_a = (D_1, \dots, D_g)^T$ and \mathbf{V}_a be the $g \times g$ matrix with (r, i) entry V_{ri} .

Second, consider the animals which have died from a tumor of interest. Table 2 is a contingency

table which shows tumor mortality data for interval I_j . Let $m_{ij} = N_{i,j-1}$ be the number of animals in group i surviving at the beginning of the interval, and $x_{ij} = \hat{d}_{ij}$ be the number of these animals dying from the tumor of interest in that interval. A vector \mathbf{D}_b that has the differences of observed and expected values using the data in Table 2 is calculated in the same way as for the incidental tumors, and the corresponding covariance matrix \mathbf{V}_b is computed.

The analysis of data on occult tumors using contexts of observation is based on the vector $\mathbf{D} = \mathbf{D}_a + \mathbf{D}_b$, with covariance matrix $\mathbf{V} = \mathbf{V}_a + \mathbf{V}_b$. Then, a dose-related trend test in the incidence of tumor of interest among dose groups can be considered by using

$$X_R = \mathbf{l}^T \mathbf{D} / \sqrt{\mathbf{l}^T \mathbf{V} \mathbf{l}}, \quad (4)$$

where $\mathbf{l} = (\ell_1, \dots, \ell_g)^T$, and ℓ_i stands for the dose metric for the i th group. Under the null hypothesis, X_R is asymptotically distributed as the standard normal.

Kodell, Chen and Moore (1994) mentioned that Peto's COD test with the conventional way of intervalization may not control size properly. They showed with 1000 two-group data sets how Peto's test could substantially exceed its nominal size in the presence of differential competing risks. However, when the incidental part of Peto's test was implemented using the NTP (National Toxicology Program) intervals (Bailer and Portier, 1988), the test maintained nominal size.

4 Simulation Study

4.1 Design of Monte Carlo Simulation Study

A Monte Carlo simulation study was conducted to evaluate the performance of the proposed testing procedure. Our modified Peto test is compared with Peto's COD test, the Poly-3 test (Bieler and Williams, 1993) and CA test. The Poly-3 test is chosen to be compared with our test because it is an age-adjusted test for data without COD information. It was recently adopted by the NTP. It is well-known that an age-adjusted test is preferable, but the CA test, which is not an age-adjusted test, is included in our simulation study as a baseline.

In the simulation, a bioassay design with four dose groups in which there were 50 animals in each group was considered. All the remaining live animals were sacrificed at the end of the experiment. The proposed testing procedure was simulated to have the NTP intervals with time points 52, 78

and 92 weeks and a terminal sacrifice at 104 weeks for both the incidental and fatal tumors. Peto's COD test using intervals defined by each animal's actual death time for the fatal tumor and the NTP intervals for the incidental tumor was also simulated to see how the probability of a Type I error was affected with correct and erroneous COD information.

Misclassification of COD was simulated as the following. Several probabilities of misclassification of COD were selected. We assumed $x\%$ chance of misclassifying an incidental tumor as fatal, and $y\%$ chance of misclassifying a fatal tumor as incidental. For the intermediate (approximately 35% of the tumors are lethal) and low (approximately 5% of the tumors are lethal) lethal tumors, the values of the (x, y) pair were chosen to be (30, 10), (50, 15) and (70, 20). For the highly lethal tumors (approximately 90% of the tumors are lethal), the error rates were chosen to be (10, 30), (15, 50) and (20, 70). The reason for choosing the opposite error rates for the highly lethal tumors was due to the fact that there were more animals which were dead from the tumor of interest than the animals which were dead from incidental causes in the case of highly lethal tumors.

It was assumed that three independent random variables completely determined the observed outcome for each animal. The random variables were the time to onset of tumor, T_1 , the time after onset until death from the tumor, T_2 , and the time to death from a competing risk, X_C . Note that $T_1 + T_2 = T_D$, where T_D represents the overall time to death from the tumor of interest. Thus the tumor of interest was present in an animal at the time of death if $T_1 \leq \min\{X_C, X_S\}$, where X_S denotes an animal's scheduled sacrifice time. An animal died from the tumor of interest if $T_D \leq \min\{X_C, X_S\}$. Otherwise, it died from a competing risk including sacrifice. Distributions of the random variables were of the form used by Kodell and Ahn (1997).

The distribution of time to onset of tumor, T_1 , was modeled by the survival function $S(t) = \exp[-\theta\delta_1(t/104)^{\delta_2}]$, where $\delta_1 \geq 0$, $\delta_2 \geq 0$, $\theta = e^\ell \geq 1$, and ℓ is a dose level. As in the NTP studies, the value of δ_2 was set to be 1.5, 3 and 6. For size evaluations (i.e., no dose effect), the probability of tumor onset by 104 weeks was chosen to be 0.05 (rare tumor), 0.15 or 0.3 (common tumor). For power comparisons, the probability of tumor onset in the highest dose group ℓ_H by 104 weeks would be x times that in control ($\ell = \ell_0$), where x is 5, 3 and 2 for background tumor rates 0.05, 0.15 and 0.3, respectively. The low and intermediate doses were chosen as $\ell_H/4$ and $\ell_H/2$, respectively. The survival rates with low tumor lethality (approximately 5% of observed tumors are COD), intermediate tumor lethality (approximately 35% of observed tumors are COD), and high

tumor lethality (approximately 90% of observed tumors are COD) were considered. The competing risks survival rate (CRSR) of either 0.5 for all the dose groups or 0.5, 0.4, 0.3 and 0.2 for control, low, intermediate, and high dose groups, respectively, were considered. These rates represent those recently observed in actual experiments. Historical data show that the survival rates of mice and rats have been declining.

In the simulation, the size (probability of a Type I error) evaluations were carried out with respect to a nominal significance level of 5%. Five thousand simulated data sets with Weibull tumor onset distributions ($\delta_2 = 1.5, 3$ or 6) for the power and the size were generated for each dose group according to each combination of three tumor onset probabilities at 104 weeks, two sets of CRSR, and three tumor lethality rates. The power evaluation also used a nominal 5% significance level for the same configurations used in the size evaluations. Thus, a total of 18 configurations for size and 18 configurations for power were considered for each model of the tumor onset distribution. The robustness of the proposed testing procedure was investigated in conjunction with the misclassification of COD in the Peto test and the performance of the Poly-3 test and the CA test. The configurations with the same CRSR in all dose groups are not reported because there was not much difference in size and power among the different tests and the misclassification of COD's.

4.2 Simulation Results

Tables 3, 4 and 5 show a size and power comparison of the tests for Weibull tumor onset data with $\delta_2 = 1.5$, $\delta_2 = 3.0$ and $\delta_2 = 6.0$, respectively, with a single terminal sacrifice. The 95% confidence interval for the true size is (4.4%, 5.6%) based on 5000 data sets.

The simulation results in Tables 3-5 show that the proposed testing procedure controlled the probability of a Type I error (size) quite well in most of the configurations. For low and intermediate lethal tumors with different CRSR, Peto's test with erroneous COD showed a substantial increase in the Type I error rate, especially for common tumors. For highly lethal tumors, Peto's test with the erroneous COD showed conservatism and a substantial loss of power. According to these results, the effect of misclassification of COD assigned by pathologists can be predicted. The new testing approach might be more reliable than Peto's test with pathologist-assigned COD information in terms of controlling size.

Our modified Peto test performed substantially better than the Poly-3 test for the data with

Weibull tumor onset distribution with power 1.5 in Table 3. The Poly-3 test showed anticonservatism in our simulation in Table 3. The reason was that the actual tumor onset distribution is Weibull with the power 1.5, but the test statistic is based on Weibull with power 3. For the cases of high tumor rate, the size of our modified Peto test was conservative for Weibull tumor onset data with power 3 and 6 for highly lethal tumors (see Tables 4 and 5). For the cases of lower tumor rate or intermediate lethality in Tables 4 and 5, our test procedure performed well. The Poly-3 trend test controlled the probability of a Type I error quite well for Weibull tumor onset data with $\delta_2 = 3$. This is an ideal situation for the Poly-3 test. These results confirm the simulation results by Kodell et al. (1994). Note that size and power of the Poly-3 test and the CA test do not change as the tumor lethality rate changes. This is because these tests do not depend on tumor lethality or COD information. The CA test was conservative for different CRSR. This result is due to the fact that the CA test is not an age-adjusted test. This behavior of the CA test is well known, and it is a major reason why age-adjusted tests have been developed. The results for the CA test are reported in this paper as baseline results for the various testing methods.

Overall, the modified Peto test controlled size slightly better than the Poly-3 test for various degrees of tumor lethality. The modified Peto test maintained desirable power for most of the configurations. In Table 3, the size of the modified Peto test was outside the 95% confidence interval in 4 of these 9 cases, whereas the size of the Poly-3 test was outside the 95% confidence interval in all 9 configurations. The misclassified COD assignment by pathologists might give either a substantial power loss or substantial anti-conservatism, depending on the lethality of the tumor. Our simulation study showed that the modified Peto test was robust for single-sacrifice data in terms of controlling the probability of a Type I error compared to other tests.

McKnight (1985) and Selwyn, Roth and Weeks (1985) commented on the difficulty of controlling the size when the death distributions are widely different in treatment and control groups. Kodell et al. (1994) also showed this by a Monte Carlo simulation study. In our simulation study, although not reported, size for the same CRSR across dose groups was close to the nominal level for all the configurations regardless of the tests considered, COD error rates of Peto's test, and tumor onset distributions. However, the size was more difficult to control if the CRSR was different among the dose groups. These results confirm the findings of the previous works.

Our modified Peto test is expected to follow an asymptotic normal distribution if there is no bias

in the attribution of tumor lethality. As seen in our simulation study, our test shows a comparable performance with the Peto test having a slight COD error. Figure 2 shows normal probability plots for the Peto test and the modified Peto test. The purpose of studying this figure is twofold. One is to examine the asymptotic normality of our test through the normal probability plots. The other is to conduct a side-by-side comparison of our test statistic with the Peto test statistic without COD error which is proven to be asymptotically normal. Among 27 configurations in the simulation, 24 of them were selected randomly to draw normal probability plots. The asymptotic normality of the proposed test could be justified in most of the cases for tumor rate 0.15 and above. For cases of tumor rate 0.05, the normal probability plots were slightly off the straight line at the tails due to the insufficient number of tumors. Exact statistical methods are commonly used for analysis of data with such a rare tumor (see Mancuso et al., 2002). Even in these cases, the proposed test appears to be better than the Peto test which is proven to follow an asymptotic normal distribution.

5 Example: ED₀₁ Data

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 (ED₀₁ study; Littlefield et al., 1980). Data in Table 6 are from groups of animals that were dosed continuously at concentrations of 0, 35, 75 or 150 ppm 2-AAF until the terminal sacrifice at 726 days. The tumors of interest were hepatocellular adenomas or carcinomas of the liver. The animals are from only one room out of 3 rooms that had a single terminal sacrifice at 24 months (room 141), and represent only animals that were designated to be sacrificed at 24 months. That is, these are 24-month cohorts.

As illustrated in Section 4.1, the NTP intervals (0-52, 52-78, 78-92 and 92-104 weeks) were used in the proposed testing procedure outlined in Section 3. In Peto's test, the NTP intervals were used for the incidental tumor analysis, and the data-determined intervals defined by animal's actual death time were used for the fatal tumor analysis. Our modified Peto test was compared with Peto's test using the COD assigned by pathologists, the Poly-3 and the CA tests. Table 7 shows the one-sided trend test for all dose groups and pairwise test results for the control and 35 ppm groups. The tumor-death-ratio and the tumor-sacrifice-ratio defined in Section 2 were 9.11% and 11.98%, respectively. It indicates that the tumor is not highly lethal. On the average, the

numbers of fatal tumors assigned by pathologists were higher than those imputed by the proposed method. This supports the argument of Kodell et al. (1995) and Ahn et al. (2000) that pathologists tend to assign too many tumors as fatal.

All the trend tests showed a highly significant dose-related trend. The results from the proposed test showed consistency with those from the other tests. The modified Peto test using the imputed COD and Peto's test using the COD assigned by pathologists gave the same conclusion. For the pairwise test, the Z -values were near the border line. Peto's COD test showed a significant dose effect, but our modified Peto test, the Poly-3 and the CA tests showed no significant dose effect at 35 ppm with a 5% significance level. This is due to a different COD assignment by pathologists and by our imputed COD. The proposed testing procedure was successfully conducted by attributing tumor lethality in the absence of COD information.

Figure 3 compares the time-to-onset survival functions obtained from the NPMLE method discussed in Section 2 using the NTP intervals. These functions and the number of fatal tumors \hat{d}_j used in the modified Peto test were obtained from the same maximum likelihood estimators. The figure also supports the test results because it shows a difference in the curves among the four dose groups.

6 Discussion

In many animal bioassays for carcinogenicity, COD is not available. Even if COD is available, it is subject to error, potentially leading to misrepresentative estimates and tests of tumor rates. The present paper developed procedures for testing dose-related trend in the absence of COD.

Our modified Peto's test was implemented by attributing tumor lethality when COD is not available. The proposed testing method gave results close to Peto's COD test with small COD assignment error and it controlled size better than the Peto test with substantial COD error. From the simulation, the proposed test procedure was found to control the probability of a Type I error reasonably well in conjunction with the attribution of tumor lethality via a constrained Newton-based NPMLE method for the data with a single sacrifice. According to our study, the Peto test was highly sensitive to the accuracy of the COD assignment. When the error of the COD assignment was severe, Peto's COD test failed to control the probability of a Type I error in the cases of low

and intermediate lethal tumors, and it experienced a substantial power loss in the case of highly lethal tumors with different CRSR.

We also compared the proposed test with the Poly-3 test because they are both age-adjusted trend tests that do not require COD information in the data. The Poly-3 test performed well and controlled the probability of a Type I error well for tumor onset data from a Weibull distribution with power 3, which is the ideal case for the test. However, the proposed procedure appeared to be slightly less model-dependent than the Poly-3 test. The proposed testing method controlled the probability of a Type I error better than the Poly-3 test for the Weibull tumor onset data with power 1.5. Asymptotic normality of the proposed test under the null hypothesis was examined by using the normal probability plots and by a comparison with the Peto test statistic with no COD which is proven to follow an asymptotic normal distribution.

The proposed testing procedure was applied to ED_{01} data to find the effect of feeding 2-AAF to female BALB/C mice, and it showed a highly significant dose-related trend for liver tumors. The results from Peto's test and the Poly-3 test were close. These results are seen in Table 7. In a pairwise test, Peto's test gave a slightly different result from our modified Peto test and the Poly-3 test.

The new testing procedure enhances the statistical analysis and interpretation of rodent tumorigenicity studies. This method may reduce the subjective bias that results from uncertainty in COD assignment, and could relieve the need for case-by-case assignment of COD or context of observation by pathologists. Interim sacrifice data as well as COD data add expense to rodent bioassays, so neither type of information is always available in the standard 2-year rodent study. The proposed testing method is a suitable statistical procedure for testing in studies that have a single terminal sacrifice.

Acknowledgements

Hojin Moon and Hongshik Ahn's work was supported by NIH grant 1 R29 CA77289-04. Hongshik Ahn's research was partially supported by the Faculty Research Participation Program at the National Center for Toxicological Research administered by the Oak Ridge Institute for Science and Education through an interagency agreement between USDOE and USFDA. The authors wish

to thank Jungnam Joo for assistance in running the Monte Carlo simulation study.

References

- Ahn, H. and Kodell, R. L., 1998: Analysis of long-term carcinogenicity studies. In: *Design and analysis of animal studies in pharmaceutical development*, Chow, S. C. and Liu, J. P. (Eds.), 259-289. Marcel Dekker, Inc., New York.
- Ahn, H., Kodell, R. L., and Moon, H., 2000: Attribution of tumor lethality and estimation of time to onset of occult tumors in the absence of cause-of-death information. *Applied Statistics* **49**, 157-169.
- Ahn, H., Moon, H., Kim, S., and Kodell, R. L., 2002: A Newton-based approach for attributing tumor lethality in animal carcinogenicity studies. *Computational statistics and Data Analysis* **38**, 263-283.
- Archer, L. E. and Ryan, L. M., 1989: On the role of cause-of-death data in the analysis of rodent tumorigenicity experiments. *Applied Statistics* **38**, 81-93.
- Armitage, P., 1955: Tests for linear trends in proportions and frequencies. *Biometrics* **11**, 375-386.
- Bailer, A. J. and Portier, C. J., 1988: Effects of treatment-induced mortality and tumor-induced mortality on tests for carcinogenicity in small samples. *Biometrics* **44**, 417-431.
- Bieler, G. S. and Williams, R. L., 1993: Ratio estimates, the delta method, and quantal response tests for increased carcinogenicity. *Biometrics* **49**, 793-801.
- Cochran, W. G., 1954: Some methods for strengthening the common χ^2 tests. *Biometrics* **10**, 417-451.
- Dewanji, A. and Kalbfleisch, J. D., 1986: Nonparametric Methods for Survival/Sacrifice Experiments. *Biometrics* **42**, 325-341.
- Gómez, G., Van Ryzin, J., 1992: Estimation of the subsurvival function for time-to-tumor in survival/sacrifice experiments. *Statistics and Probability Letters* **13**, 5-13.
- Kodell, R. L. and Ahn, H., 1997: An age-adjusted trend test for the tumor incidence rate. *Biometrics* **53**, 1467-1474.
- Kodell, R. L. and Ahn, H., Pearce, B. A., and Turturro, A., 1997: Age-adjusted trend test for the tumor incidence rate. *Drug Information Journal* **31**, 471-487.
- Kodell, R. L., Blackwell, B.-N., Bucci, T. J., and Greenman, D. L., 1995: Cause-of-death assignment at the national center for toxicological research. *Toxicologic Pathology* **23**, 241-247.
- Kodell, R. L. and Chen, J. J., 1987: Handling cause of death in equivocal cases using the EM algorithm (with discussion). *Communications in Statistics - Theory and Methods* **16**, 2565-2585.

- Kodell, R. L., Chen, J. J., and Moore, G. E., 1994: Comparing distributions of time to onset of disease in animal tumorigenicity experiments. *Communications in Statistics - Theory and Methods* **23**, 959-980.
- Kodell, R. L. and Nelson, C. J., 1980: An illness-death model for the study of the carcinogenic process using survival/sacrifice data. *Biometrics* **36**, 267-277.
- Kodell, R. L., Shaw, G. W., and Johnson, A. M., 1982: Nonparametric joint estimators for disease resistance and survival functions in survival/sacrifice experiments. *Biometrics* **38**, 43-58.
- Lagakos, S. W., 1982: An evaluation of some two-sample tests used to analyze animal carcinogenicity experiments. *Utilitas Mathematica* **21B**, 239-260.
- Lagakos, S. W. and Louis, T. A., 1988: Use of tumour lethality to interpret tumorigenicity experiments lacking cause-of-death data. *Applied Statistics* **37**, 169-179.
- Lindsey, J. C. and Ryan, L. M., 1994: A comparison of continuous- and discrete-time three-state models for rodent tumorigenicity experiments. *Environmental Health Perspectives* **102**(Suppl. 1), 9-17.
- Littlefield, N. A., Farmer, J. H., Gaylor, D. W., and Sheldon, W. G., 1980: Effects of dose and time in a long-term, low-dose carcinogenic study. *Journal of Environmental Pathology and Toxicology* **3**, 17-34.
- Malani, H. M. and Van Ryzin, J., 1988: Comparison of two treatments in animal carcinogenicity experiments. *Journal of the American Statistical Association* **83**, 1171-1177.
- Mancuso, J. Y., Ahn, H., Chen, J. J. and Mancuso, J. P., 2002: Age-adjusted exact trend tests in the event of rare occurrences. *Biometrics*. To appear.
- McKnight, B., 1985: Discussion of papers on statistical tests for carcinogenic effects. *Proceedings of the Symposium on Long-Term Animal Carcinogenicity Studies: A Statistical Perspective*, 107-111.
- McKnight, B. and Crowley, J., 1984: Tests for differences in tumor incidence based on animal carcinogenesis experiments. *Journal of the American Statistical Association* **79**, 639-648.
- Moon, H., Ahn, H., Kodell, R. L., and Pearce, B. A., 1999: A comparison of a mixture likelihood method and the EM algorithm for an estimation problem in animal carcinogenicity studies. *Computational Statistics and Data Analysis* **31**, 227-238.
- Peto, R., 1974: Guidelines on the analysis of tumour rates and death rates in experimental animals. *British Journal of Cancer* **29**, 101-105.
- Peto, R., Pike, M. C., Day, N. E., Gray, R. G., Lee, P. N., Parish, S., Peto, J., Richards, S., and Wahrendorf, J., 1980: Guidelines for simple, sensitive significance tests for carcinogenic effects in long-term animal experiments. Annex to: *Long-term and Short-term Screening Assays for Carcinogens: a Critical Appraisal*. IARC Monographs, Supplement 2, 311-426. International Agency for Research on Cancer: Lyon, France.

- Racine-Poon, A. and Hoel, D. G., 1984: Nonparametric estimation of the survival function when cause of death is uncertain. *Biometrics* **40**, 1151-1158.
- Selwyn, M. R., Roth, A. J., and Weeks, B. J., 1985: The weighted prevalence method for analyzing nonlethal tumor data. *Proceedings of the Symposium on Long-Term Animal Carcinogenicity Studies: A Statistical Perspective*, 85-90.
- Turnbull, B. W. and Mitchell, T. J., 1984: Nonparametric estimation of the distribution of time to onset for specific diseases in survival/sacrifice experiments. *Biometrics* **40**, 41-50.

Table 1: Tumor prevalence data for interval I_j .

	Dose group				Total
	1	2	...	g	
#with tumors	y_{1j}	y_{2j}	...	y_{gj}	$y_{.j}$
#without tumors	$n_{1j} - y_{1j}$	$n_{2j} - y_{2j}$...	$n_{gj} - y_{gj}$	$n_{.j} - y_{.j}$
#deaths from competing causes	n_{1j}	n_{2j}	...	n_{gj}	$n_{.j}$

Table 2: Tumor mortality data for interval I_j .

	Dose group				Total
	1	2	...	g	
#fatal tumor deaths in I_j	x_{1j}	x_{2j}	...	x_{gj}	$x_{.j}$
	$m_{1j} - x_{1j}$	$m_{2j} - x_{2j}$...	$m_{gj} - x_{gj}$	
#surviving in I_j	m_{1j}	m_{2j}	...	m_{gj}	$m_{.j}$

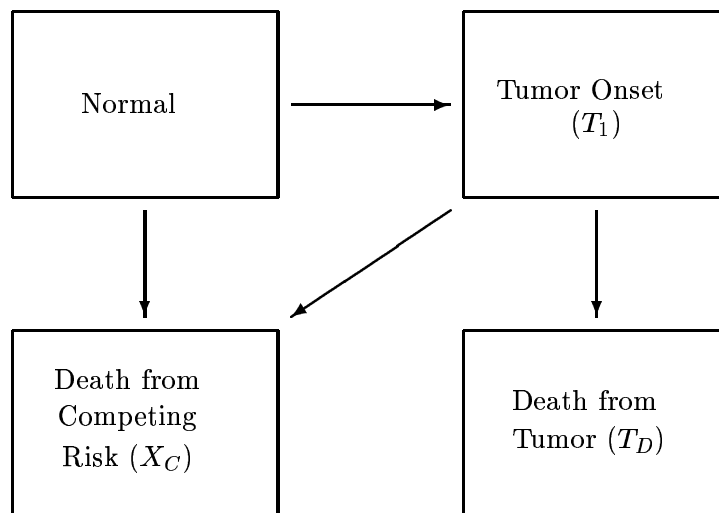


Figure 1: Illustration of illness and death, with possible transitions, in rodent bioassay from Kodell and Nelson (1980).

Table 3: Simulated size and power (%) corresponding to nominal 5% significance level for trend tests applied to single-sacrifice data with a Weibull tumor onset distribution ($\delta_2 = 1.5$). All tests are based on a dose scaling of 0, 1, 2, 4 with 5000 trials for each configuration and on the different CRSR^a.

	Tumor Onset Prob. ^b	Tumor Lethal. Prob. ^c	Peto with imputed COD	Peto with correct COD	Peto with Wrong COD			Poly-3	CA ^d
					30-10 ^e	50-15	70-20		
Size	.05	≈ .05	6.6	4.9	5.5	6.6	7.3	6.9	4.1
	.15	≈ .05	5.6	4.8	6.2	9.3	10.4	6.7	2.9
	.30	≈ .05	5.6	4.6	7.9	12.4	15.5	6.5	2.0
Power	.05	≈ .05	89.4	87.3	90.0	92.3	93.5	91.6	83.7
	.15	≈ .05	95.1	92.6	95.8	97.8	98.5	96.2	89.0
	.30	≈ .05	91.6	88.2	94.0	97.3	98.1	92.9	80.1
Size	.05	≈ .35	5.9	5.4	5.9	6.4	6.4	6.9	4.1
	.15	≈ .35	4.7	4.8	5.8	6.9	8.1	6.7	2.9
	.30	≈ .35	3.9	4.6	6.2	8.6	10.1	6.5	2.0
Power	.05	≈ .35	88.2	86.6	88.3	90.1	90.8	91.6	83.7
	.15	≈ .35	92.9	92.6	94.6	96.3	96.7	96.2	89.0
	.30	≈ .35	86.5	86.6	91.3	94.5	95.6	92.9	80.1
Size	.05	≈ .90	5.5	4.8	3.0	2.3	1.9	6.9	4.1
	.15	≈ .90	4.7	5.0	2.4	1.2	0.8	6.7	2.9
	.30	≈ .90	3.6	4.6	2.0	1.0	0.5	6.5	2.0
Power	.05	≈ .90	88.7	88.2	79.9	68.8	61.5	91.6	83.7
	.15	≈ .90	93.7	93.1	85.4	71.4	61.0	96.2	89.0
	.30	≈ .90	86.4	89.5	76.5	57.0	45.6	92.9	80.1

^a 0.5, 0.4, 0.3, 0.2 for control, low, intermediate and high dose groups, respectively.

^b Cumulative tumor onset probability (in control for power) at 104 weeks in absence of competing risks.

^c Proportion of observed tumors that actually result in death. Same probability in all dose groups.

^d Cochran-Armitage test.

^e $x - y$ represents $x\%$ misclassified an incidental tumor as fatal, $y\%$ misclassified a fatal tumor as incidental.

Table 4: Simulated size and power (%) corresponding to nominal 5% significance level for trend tests applied to single-sacrifice data with a Weibull tumor onset distribution ($\delta_2 = 3.0$). All tests are based on a dose scaling of 0, 1, 2, 4 with 5000 trials for each configuration and on the different CRSR^a.

	Tumor Onset Prob. ^b	Tumor Lethal. Prob. ^c	Peto with imputed COD	Peto with correct COD	Peto with Wrong COD			Poly-3	CA ^d
					30-10 ^e	50-15	70-20		
Size	.05	≈ .05	5.9	5.2	5.6	6.0	6.6	5.9	3.5
	.15	≈ .05	4.7	4.2	5.3	7.2	8.1	4.8	2.2
	.30	≈ .05	4.5	4.1	6.7	9.7	11.7	4.0	1.1
Power	.05	≈ .05	84.2	82.0	84.3	86.8	88.0	85.4	75.0
	.15	≈ .05	90.5	88.9	92.3	94.9	95.9	90.9	79.3
	.30	≈ .05	85.4	83.0	89.8	94.0	95.2	84.4	63.9
Size	.05	≈ .35	5.2	4.8	5.0	5.3	5.5	5.9	3.5
	.15	≈ .35	3.8	4.3	5.1	5.7	6.4	4.8	2.2
	.30	≈ .35	2.6	4.8	5.8	7.6	8.5	4.0	1.1
Power	.05	≈ .35	81.6	82.8	84.4	86.0	86.8	85.4	75.0
	.15	≈ .35	86.1	89.4	91.3	93.0	93.5	90.9	79.3
	.30	≈ .35	76.4	82.7	86.8	89.9	91.6	84.4	63.9
Size	.05	≈ .90	5.0	4.5	3.3	2.5	1.9	5.9	3.5
	.15	≈ .90	4.1	4.7	2.9	1.5	1.1	4.8	2.2
	.30	≈ .90	2.4	4.8	2.0	0.9	0.6	4.0	1.1
Power	.05	≈ .90	83.2	83.2	75.3	64.9	58.4	85.4	75.0
	.15	≈ .90	87.3	91.4	82.5	68.8	60.6	90.9	79.3
	.30	≈ .90	73.7	86.7	72.1	53.1	42.5	84.4	63.9

^a 0.5, 0.4, 0.3, 0.2 for control, low, intermediate and high dose groups, respectively.

^b Cumulative tumor onset probability (in control for power) at 104 weeks in absence of competing risks.

^c Proportion of observed tumors that actually result in death. Same probability in all dose groups.

^d Cochran-Armitage test.

^e $x - y$ represents $x\%$ misclassified an incidental tumor as fatal, $y\%$ misclassified a fatal tumor as incidental.

Table 5: Simulated size and power (%) corresponding to nominal 5% significance level for trend tests applied to single-sacrifice data with a Weibull tumor onset distribution ($\delta_2 = 6.0$). All tests are based on a dose scaling of 0, 1, 2, 4 with 5000 trials for each configuration and on the different CRSR^a.

	Tumor Onset Prob. ^b	Tumor Lethal. Prob. ^c	Peto with imputed COD	Peto with correct COD	Peto with Wrong COD			Poly-3	CA ^d
					30-10 ^e	50-15	70-20		
Size	.05	≈ .05	5.3	5.3	5.2	5.6	5.8	4.7	2.5
	.15	≈ .05	4.2	4.2	4.9	6.0	6.4	3.4	1.3
	.30	≈ .05	3.5	3.2	4.4	6.3	6.9	2.2	0.6
Power	.05	≈ .05	75.9	75.3	76.8	79.0	80.4	73.9	60.5
	.15	≈ .05	82.0	81.6	85.1	87.9	89.4	79.0	60.2
	.30	≈ .05	74.0	72.0	79.3	84.9	87.3	66.3	39.2
Size	.05	≈ .35	4.6	4.0	4.2	4.5	4.5	4.7	2.5
	.15	≈ .35	3.0	4.3	4.7	5.0	5.4	3.4	1.3
	.30	≈ .35	2.1	3.5	4.3	5.0	5.6	2.2	0.6
Power	.05	≈ .35	72.3	73.7	74.7	76.0	76.6	73.9	60.5
	.15	≈ .35	75.7	82.1	83.7	85.7	86.7	79.0	60.2
	.30	≈ .35	63.4	73.1	76.7	80.4	82.1	66.3	39.2
Size	.05	≈ .90	4.6	4.4	3.3	2.6	2.1	4.7	2.5
	.15	≈ .90	3.3	4.1	2.5	1.5	1.0	3.4	1.3
	.30	≈ .90	1.7	4.7	2.4	1.1	0.7	2.2	0.6
Power	.05	≈ .90	74.1	76.1	69.3	61.4	57.3	73.9	60.5
	.15	≈ .90	76.0	86.1	76.8	65.0	57.7	79.0	60.2
	.30	≈ .90	55.6	80.1	66.7	49.6	41.0	66.3	39.2

^a0.5, 0.4, 0.3, 0.2 for control, low, intermediate and high dose groups, respectively.

^bCumulative tumor onset probability (in control for power) at 104 weeks in absence of competing risks.

^cProportion of observed tumors that actually result in death. Same probability in all dose groups.

^dCochran-Armitage test.

^e $x - y$ represents $x\%$ misclassified an incidental tumor as fatal, $y\%$ misclassified a fatal tumor as incidental.

Table 6: Frequency data from ED₀₁ study.

Dose (ppm)	j^a	\tilde{d}_j^b	\tilde{d}_j^c	$a_{1j} + d_j$	b_{1j}	a_{2j}	b_{2j}
0	1	0	0	0	9	0	0
	2	0	0	0	15	0	0
	3	0	1	1	34	0	0
	4	0	1	1	60	7	137
35	1	0	0	1	9	0	0
	2	1	3	3	31	0	0
	3	0	1	1	55	0	0
	4	0	0	1	80	18	192
75	1	0	0	0	10	0	0
	2	0	0	1	16	0	0
	3	1	2	3	28	0	0
	4	1	3	10	41	22	101
150	1	0	0	0	5	0	0
	2	0	1	2	14	0	0
	3	1	2	6	24	0	0
	4	5	7	15	18	16	33

^aTime intervals 1-4 represent, respectively, 0-364, 365-546, 547-644, 645-726 days.

^bImputed number of fatal tumors obtained by the proposed Newton-based NPMLE method.

^cNumber of fatal tumors assigned by pathologists.

Table 7: Test results for proposed modified Peto test with imputed COD, Peto's test with COD assigned by pathologists, the Poly-3 trend test and the CA test for the ED₀₁ data given in Table 6. All tests are one-sided.

			Proposed	Peto	Poly-3	CA
Incidental	Z value		8.27	8.27		
	p -value		≈ 0	≈ 0		
Trend test	Fatal	Z value	4.83	4.50		
		p -value	≈ 0	≈ 0		
Combined	Z value		9.34	9.41	9.40	8.69
	p -value		≈ 0	≈ 0	≈ 0	≈ 0
Incidental	Z value		1.49	1.68		
	p -value		0.068	0.046		
Pairwise test for control & 35 ppm group	Fatal	Z value	0.82	0.55		
		p -value	0.206	0.291		
Combined	Z value		1.61	1.75	1.58	1.57
	p -value		0.054	0.040	0.057	0.058

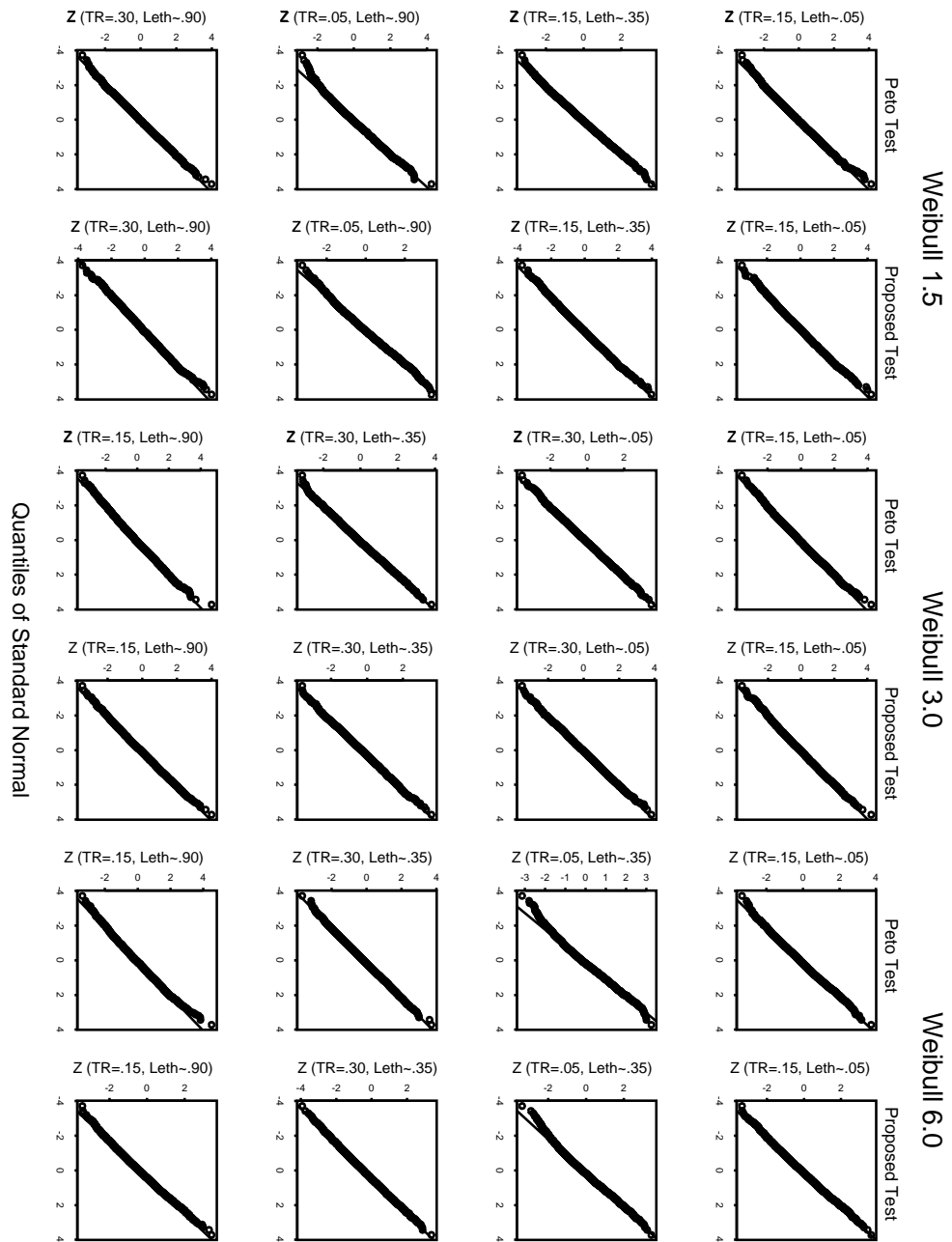


Figure 2: Normal probability plots obtained from the Peto test and the proposed test

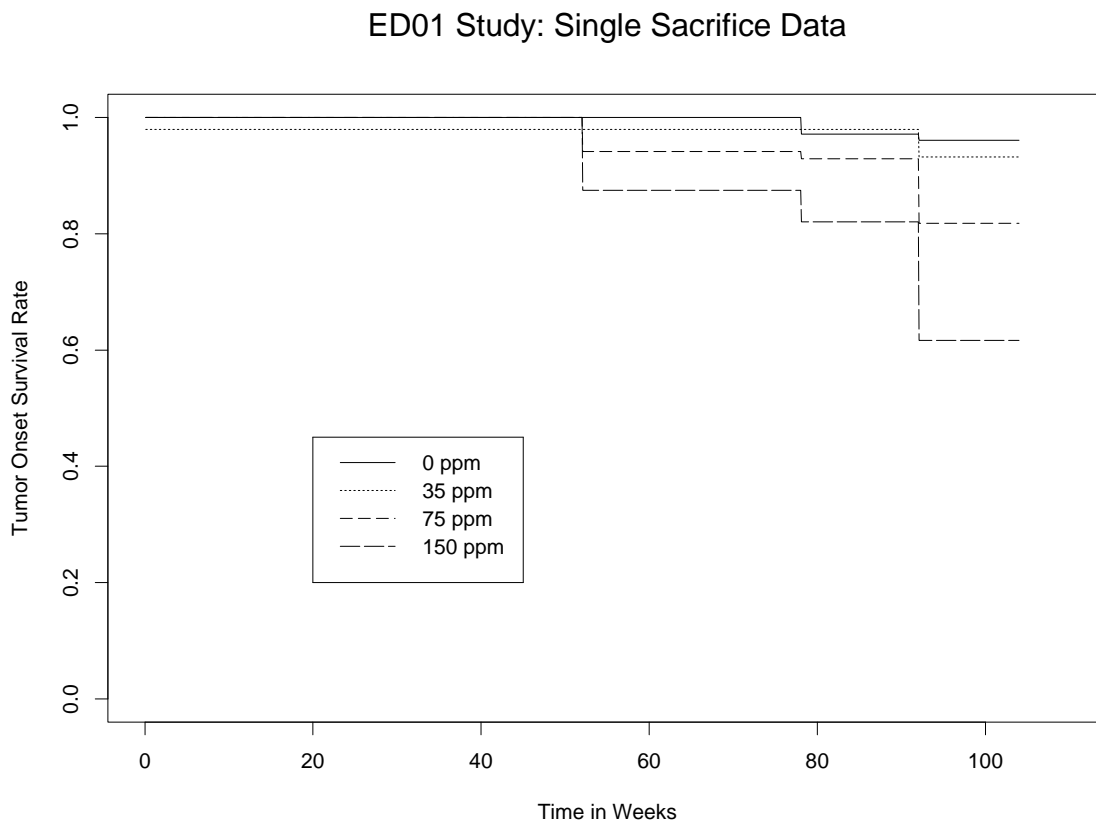


Figure 3: Time-to-tumor-onset survival functions obtained from the proposed Newton-based NPMLE estimation method for ED_{01} data.