

A Weight-Adjusted Peto's Test When Cause of Death is Not Assigned

Hojin Moon¹, Hongshik Ahn², J. Jack Lee¹ and Ralph L. Kodell³

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

²Department of Applied Mathematics and Statistics, State University of New York
Stony Brook, NY 11794-3600

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

Abstract

A new statistical testing approach using a weighted logrank statistic is developed for rodent tumorigenicity assays that have a single terminal sacrifice but not cause-of-death data. Instead of using cause-of-death assignment by pathologists, the number of fatal tumors is estimated by a constrained nonparametric maximum likelihood estimation method. For data lacking cause-of-death information, the Peto test is modified with estimated numbers of fatal tumors and a Fleming-Harrington-type weight, which is based on an estimated tumor survival function. A bootstrap resampling method is used to estimate the weight function. The proposed testing method with the weight adjustment appears to improve the performance in various situations of single-sacrifice animal experiments. A Monte Carlo simulation study for the proposed test is conducted to assess size and power of the test. This testing approach is illustrated using a real data set.

Key Words: Bioassay; Bootstrap; Competing risk; Fatal tumor; Single sacrifice; Weight function.

1 Introduction

Peto's test (Peto et al., 1980), recommended by the International Agency for Research on Cancer (IARC), is a widely used statistical test for a dose-related trend in the occurrence of tumors. It requires an assessment of the context of observation for each tumor or an assignment of cause of death (COD) for each animal. However, this information is not always available and even when it is available, the accuracy of the assignment by pathologists is subject to error (Racine-Poon and Hoel, 1984; Lagakos and Louis, 1988; Kodell et al., 1995). The National Toxicology Program (NTP) has been analyzing data lacking COD using other statistical methods.

Peto's test is highly sensitive to an error of COD assignment (see Moon et al., 2002). Recently, Ahn et al. (2000) developed a method of statistically imputing the lethality of a tumor in order to reduce the possible bias of the COD assignment by pathologists and to enable calculation of the Peto test even in the absence of COD information. Moon et al. (2002) further modified this estimation method and applied these imputed numbers of deaths from the tumor of interest and numbers of deaths with incidental tumor to a modification of Peto's test. To further improve the performance of the test, we propose a Fleming-Harrington-type (Fleming and Harrington, 1981) weight-adjusted Peto's test via a bootstrap resampling method, in addition to the modification by Moon et al. The size of this test becomes closer to the nominal significance level according to our simulation study.

Fleming and Harrington (1981) proposed a general class of tests, including a logrank test and a version of the Gehan-Wilcoxon test. Their weight function is $W(t_j) = [\hat{P}(t_{j-1})]^\beta [1 - \hat{P}(t_{j-1})]^\gamma$, where $\beta \geq 0, \gamma \geq 0$ and $\hat{P}(t)$ is the product-limit estimator for $t_{j-1} \leq t < t_j$. When $\beta = 0$ and $\gamma = 0$, it becomes the logrank test. When $\beta = 1$ and $\gamma = 0$, it is very close to a modification of the Gehan-Wilcoxon test. When $\beta > \gamma$, this test assigns more weight on the early failure times. On the other hand, the test gives more weight to the late failure times when $\beta < \gamma$. An adverse effect was noticed in the evaluation of the size of the proposed test in a pre-simulation study when

the weight with either $\beta > \gamma$ or $\beta < \gamma$ was considered. In this study, therefore, a weight with only one parameter $\rho(= \beta = \gamma)$ is used. We develop a method of estimating an optimal value of ρ using a bootstrap resampling method.

The current study is applied to analyze animal carcinogenicity data from the ED₀₁ study conducted at the National Center for Toxicological Research (NCTR). The experiment was conducted to test the carcinogenic effect of feeding 2-acetylaminofluorene (2-AAF) to female mice (Littlefield et al., 1980). We focus on a dose-related trend of liver tumor incidence in this study. Table 1 shows frequency data from a subset of the ED₀₁ study. These data are summarized based on standard intervals used in the NTP. The animals are from only one room (room 141) out of 3 rooms that had a 24-month sacrifice (rooms 141, 142 and 143). These are the cohorts of animals that were designated to be sacrificed at 24 months.

(Table 1 is here.)

The purpose of this paper is to develop a new statistical method using a weighted Peto test statistic for testing the dose-related trend of a tumor. Since most of the animal tumorigenicity studies are currently designed with a single sacrifice at the end of an experiment, the proposed method is aimed for rodent tumorigenicity assays that have only a terminal sacrifice but not COD data, although it can also be used for studies with interim sacrifices. The tumor survival function estimated by a modification of the NPMLE method of Ahn et al. (2000) is used in the new testing procedure. The weight-adjusted Peto test is compared with three tests: the proposed modified Peto test without weight adjustment (Moon et al., 2002), the Peto test, and the Poly-3 test (Bieler and Williams, 1993). According to our simulation study, the proposed testing method shows an improvement in various situations for experiments with a single terminal sacrifice.

2 Methods: A Modification of Peto’s Test

We implement the proposed weight-adjusted Peto test with the estimated tumor lethality and a Fleming-Harrington-type weight, a function of the estimated tumor survival function. A bootstrap resampling method is used for estimating the weight function. In Section 2.1, we briefly describe the NPMLE method of Moon et al. (2002) to estimate the tumor lethality and tumor survival functions. The proposed test is described in detail in Section 2.2.

2.1 Nonparametric Maximum Likelihood Estimation of Tumor Lethality

We consider an experiment with a control and $g - 1$ dose groups and suppose that N_i animals in the i -th treatment 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 without a tumor on day zero of the experiment. We divide the time scale into J intervals such that the j -th interval is given by $I_j = (t_{j-1}, t_j]$, $j = 1, \dots, J$. (Note that $t_0 = 0$ and t_J denotes the terminal sacrifice time point.)

We let $S(t)$ be a survival function with respect to a random variable T_1 representing time to onset of the tumor of interest, and define $P(t)$ as a survival function with respect to a random variable T_D , where T_D represents the overall time to death resulting from the tumor of interest. Similarly, we define $Q(t)$ as a survival function with respect to a random variable X_C , representing time to death from a cause other than the tumor of interest (competing risks). Ahn et al. (2000) introduced the contributions to the likelihood function in terms of $S(t)$, $P(t)$ and $Q(t)$ for multiple-sacrifice data as if complete information on the COD is available for the experimental animals by assuming that T_1 and T_D are independent of X_C . Moon et al. (2002) formulated the likelihood contributions for single-sacrifice data by relaxing the independence assumption. Instead, they introduced an imposed assumption that T_1 and T_D occur before X_C in an interval. This assumption is commonly made in practice (Malani and Van Ryzin, 1988; Kodell and Ahn, 1997). Probabilities that contribute to the likelihood function for complete data are derived as in Table 2. For experiments with a single

terminal sacrifice, a_{2j} and b_{2j} are zero, except for the last interval.

(Table 2 is here.)

After a reparameterization using $\pi_j = S(t_j)/P(t_j)$, $p_j = P(t_j)/P(t_{j-1})$ and $q_j = Q(t_j)/Q(t_{j-1})$ for all j , we derive the log-likelihood function for complete data are derived as

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

where N_j is the number of live animals at t_j , with the following inequality constraints:

$$0 \leq \pi_j, p_j \text{ and } q_j \leq 1 \text{ and } \pi_j p_j \leq \pi_{j-1}, \quad (2)$$

for $j = 1, \dots, J$. We obtain the maximum likelihood estimates of π_j , p_j and q_j for experiments lacking COD information by maximizing (1) subject to (2) via a Newton-based method proposed by Ahn et al. (2002). Among all possible pairs (d_j, a_{1j}) of integer values, given only the total number of tumor death animals $d_j + a_{1j}$, the function l is maximized for each pair of (d_j, a_{1j}) . The values of (d_j, a_{1j}) with the largest maximum likelihood and the corresponding MLE's are chosen as the desired estimates. Further details are given in Moon et al. (2002).

Given estimates $\hat{\pi}_j$, \hat{p}_j and \hat{q}_j , we estimate the survival functions $P(t)$, $Q(t)$ and $S(t)$ for $t_{j-1} < t \leq t_j$ by

$$\hat{P}(t) = \prod_{k=1}^j \hat{p}_k, \quad \hat{Q}(t) = \prod_{k=1}^j \hat{q}_k, \quad \text{and} \quad \hat{S}(t) = \hat{\pi}_j \hat{P}(t_j), \quad (3)$$

for $j = 1, \dots, J$.

2.2 Weight-Adjusted Peto's Test via a Bootstrap Resampling Method

Peto's test segregates animals according to whether or not specific tumors of interest caused the animal's death (Peto et al., 1980). In this section, we propose a weight adjustment to Peto's test with a Fleming-Harrington-type weight using estimated tumor lethality and estimated tumor

survival function via a bootstrap resampling method for data with unknown COD. This modification can be done by applying $\hat{P}(t_j)$, \hat{d}_j and \hat{a}_{1j} obtained in Section 2.1, and N_j , b_{1j} , a_{2j} and b_{2j} obtained from data for $j = 1, \dots, J$.

Consider a weight function $W(t_j) = [\hat{P}(t_j)(1 - \hat{P}(t_j))]^\rho$ for $\rho \geq 0$, where $\hat{P}(t)$ is an estimate of a survival function for $t_{j-1} < t \leq t_j$. Using the MLE's, the estimate of the tumor survival function $\hat{P}(j)$ in (3) can be obtained at the j -th interval. The weight function for fatal and incidental tumors is of the form $\bar{W}(j) = [\sum_{i=1}^g W_i(j)]/g$, where $W_i(j) = [\hat{P}_i(j)(1 - \hat{P}_i(j))]^\rho$. We propose an estimation of ρ by a bootstrap resampling method. Initially, a bootstrap sample of $\sum_{i=1}^g N_i$ animals is randomly chosen from the pooled original data with replacement regardless of the dose group. Then, the regenerated $\sum_{i=1}^g N_i$ animals are randomly reassigned into g groups. We expect that these new g groups are homogeneous. Thereafter, the test is conducted using various values of ρ on a certain grid $[0, b)$. The choice of b is illustrated in Section 3. Next, we choose ρ that gives the p -value closest to the nominal significance level α . In this study, we use $\alpha = 0.05$. We repeat the above steps for generating each bootstrap sample using the same original data set. Let ρ_i be the bootstrap estimate of ρ for the i th bootstrap sample. Then the average $\rho^* = [\sum_{i=1}^B \rho_i]/B$ is the final bootstrap estimate of ρ , where B is the number of bootstrap samples. Five thousand bootstrap samples ($B = 5000$) are used in this study.

First, we consider the animals that did not have the specific tumor before death, and tumor-bearing animals that did not die from the tumor of interest. 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 $i = 1, \dots, g$. For each interval I_j , the tumor prevalence data may be summarized in a $2 \times g$ table, as in Table 3. All tumors found in sacrificed animals are classified as incidental. The intervals defined by the pre-assigned NTP intervals (Bailer and Portier, 1988) are used to implement the incidental part of our weight-adjusted Peto's test.

(Table 3 is here.)

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 $i = 1, \dots, g$. We define D_i and V_{ri} using a weight function similar to the one in Klein and Moeschberger (1997):

$$D_i = \sum_{j=1}^m \overline{W}(j)(y_{ij} - E_{ij}) \quad \text{and} \quad V_{ri} = \sum_{j=1}^m \overline{W}(j)^2 \kappa_j K_{rj} (\delta_{ri} - K_{ij}), \quad (4)$$

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. We let $\mathbf{D}_a = (D_1, \dots, D_g)^T$ and let \mathbf{V}_a be the $g \times g$ matrix with (r, i) entry V_{ri} .

Second, we consider the animals that died with a tumor of interest. Peto's test uses each fatal death time to define an interval, in practice. On the other hand, the NTP intervals are used for fatal tumors in our weight-adjusted Peto's test to reduce the bias for the estimation of lethality. Table 4 is a contingency table for an interval I_j . We 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 estimated number of these animals dying from the tumor of interest in that interval (Moon et al., 2002). We calculate vector \mathbf{D}_b that has the differences of observed and expected values using the data in Table 4 in the same way as for the incidental tumors and also calculate the corresponding covariance matrix \mathbf{V}_b in the same way.

(Table 4 is here.)

The analysis of data on occult tumors using the context 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 can be considered by using

$$X_R = \mathbf{l}^T \mathbf{D} / \sqrt{\mathbf{l}^T \mathbf{V} \mathbf{l}}, \quad (5)$$

where $\mathbf{l} = (\ell_1, \dots, \ell_g)^T$, and ℓ_i stands for the dose metric for the i -th group. It is obtained by a

standardization using (4), and we achieve the asymptotic normality by a linear transformation of the Peto test statistic which is shown to be asymptotic normal. Therefore, under the null hypothesis, X_R is asymptotically distributed as a standard normal. We compare the proposed weight-adjusted Peto test (with the estimated tumor lethality and the estimated tumor survival function via the bootstrap resampling method) with three tests: the Poly-3 test (Bieler and Williams, 1993), a modified Peto's test without weight-adjustment (Moon et al., 2002) and Peto's test with both correct and inaccurate COD information. We choose the Poly-3 test for a comparison with our test because it is an age-adjusted test for data without COD information. We compare the proposed test with a modified Peto's test without adjustment to show the improvement of the weight-adjustment. The Peto test with inaccurate COD information is chosen to represent potential errors in COD assignment by pathologists.

3 Simulation Study

We conduct a Monte Carlo simulation study to evaluate the performance of the proposed testing procedure. In the simulation, we consider a bioassay design of 4 dose groups with 50 animals per group. At the end of the experiment, all the remaining live animals are sacrificed. For analyzing single-sacrifice data with the proposed weight-adjusted Peto test using estimated tumor lethality, we use the NTP intervals (0-52, 53-78, 79-92 and 93-104 weeks) for both the incidental and fatal tumors. We also simulate Peto's test using the NTP intervals for the incidental tumor and using intervals defined by each animal's actual death time for the fatal tumor to show the probability of a Type I error affected by correct and erroneous COD information.

The distributions of time to tumor onset, time to death from the tumor of interest, and time to death from competing risks are of the form used by Ahn et al. (2000). For the probability of tumor onset by 104 weeks, we choose 0.05 (rare tumor), 0.15 or 0.3 (common tumor) for size simulations in all 4 dose groups. For power comparisons, we choose the probability of tumor onset such that

tumor onset probability in the highest dose group by 104 weeks is 5, 3 and 2 times the background tumor rates of 0.05, 0.15 and 0.3, respectively.

We carry out size evaluations with respect to a nominal significance level of 5%. We generate 5000 simulated data sets with Weibull tumor onset distributions (with shape parameters of 1.5, 3.0 and 6.0) for power and size for each dose group according to each combination of three tumor onset probabilities (0.05, 0.15 and 0.3) at 104 weeks, two sets of competing risks survival rates (CRSR: one for 0.5 across dose groups and the other for 0.6, 0.5, 0.4 and 0.3 for control, low, intermediate and high dose groups, respectively), and three tumor lethality rates (approximately 5%, 35% and 60%). These rates represent those recently observed in actual experiments. The power evaluation also uses 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 are considered for each model of the tumor onset distribution. The robustness of the proposed testing procedure is investigated in conjunction with a practical consideration of the misclassification of COD by pathologists in Peto's test and the performance of the Poly-3 test.

The configurations with the same CRSR in all dose groups are excluded because there is not much difference in size and power among the different tests and the misclassification of COD's. This is expected due to the equal survival rate across dose groups. Moon et al. (2002) and Mancuso et al. (2002) also observe the similarity of the performance of the various tests when CRSR is the same across dose groups.

For the probabilities of misclassification of COD, we assume an $x\%$ chance of misclassifying an incidental tumor as fatal, and a $y\%$ chance of misclassifying a fatal tumor as incidental. When the tumors are of low and intermediate lethality (where approximately 5% and 35% of the tumors are lethal, respectively), we choose the values of the (x, y) pair as (30, 10), (50, 15) and (70, 20). For the highly lethal tumors (where approximately 60% of the tumors are lethal), we choose the error rates of (10, 30), (15, 50) and (20, 70).

While two parameters, β and γ , were used in the Fleming-Harrington weight function, a parameter $\rho(= \beta = \gamma)$ in Section 2.2 was chosen via the bootstrap resampling method. If more weight is given to early failure times, our test becomes conservative, which adversely affects the Weibull tumor onset distribution with a shape parameter of 6. On the other hand, if more weight is given to late failure times, our test becomes anticonservative, which adversely affects the Weibull tumor onset distribution with a shape parameter of 1.5. The weights are applied only if the tumor onset probability among all dose groups is greater than 5%, because the normal approximation of Peto's test may be unreliable when the tumor onset probability is lower than 5%. For tumors with rare occurrences, an exact test is recommended. For example, see Mancuso et al. (2002).

We simulate the weight-adjusted Peto test with the ρ estimated by the bootstrap resampling method. Figure 1 shows the simulated size of the proposed test for each value of ρ . From this figure, the range of parameter ρ in our weight function may be considered to be $[0, 1)$, i.e., $b = 1$ due to inflation of the size for the Weibull tumor onset distribution with a shape parameter of 1.5 and for 5% tumor lethality among Weibull distributions when $\rho > 1.0$. Our proposed weight-adjusted Peto's test controls the size better with ρ^* (given in Tables 5 - 7) from the bootstrap resampling method compared to the test without weight.

(Figure 1 is here.)

Table 5 shows the size and power of the trend tests applied to the Weibull distribution (with a shape parameter of 1.5) of the tumor onset data with a single sacrifice, which is based on the nominal 5% significance level for trend tests. On the basis of 5000 data sets, the 95% confidence interval for the nominal size is (4.4%, 5.6%). Our test method shows a substantial improvement for various degrees of tumor lethality. Three out of 9 cases are outside of the 95% confidence interval in the weight-adjusted Peto test. On the other hand, none of them are in the 95% confidence interval for the Poly-3 test. For the Peto test with erroneous COD, the size shows anticonservatism for large error rates in tumors of low and intermediate lethality as shown in Moon et al. (2002). The

weight-adjusted Peto's test maintained desirable power for most of the configurations. The Poly-3 trend test showed anticonservatism in our simulation as described in Moon et al. (2002). Note that the size and power of the Poly-3 test do not change as the tumor lethality rate changes because they do not depend on tumor lethality or COD information.

(Table 5 is here.)

In Tables 6 and 7, the size from the proposed test shows a nice improvement compared to the size from the modified Peto's test without weight adjustment. The proposed test seems to control size equally well with both the Peto test with correct COD and the Poly-3 test. In particular, the proposed test shows a substantial improvement compared to a modified Peto's test without adjustment ($\rho = 0$) especially for a tumor rate of 15% and higher. The weight-adjusted Peto test appears to be robust compared to other tests in terms of controlling the probability of a Type I error.

(Tables 6 and 7 are here.)

Overall, the weight-adjusted Peto test controls size better than the Poly-3 test, modified Peto's test without adjustment, and Peto's test with inaccurate COD for various degrees of tumor lethality in this simulation study. The proposed test seems to maintain a desirable power. The proposed weighted Peto statistic improves the control of size in various degrees of Weibull tumor onset distributions.

4 Example: Single-Sacrifice Data from ED₀₁ Study

The data given in Section 1 are analyzed using the proposed test, the Poly-3 test, and the modified Peto test of Moon et al. (2002). All trend tests with 8 groups are highly significant with p -value less than 0.0001. In this paper, the low dose effect of feeding 2-AAF to female mice is considered. The lowest four dose groups from the data are chosen, and the frequencies are given in Table 8. Table 9 shows the test results for the proposed test with bootstrap estimate ρ^* , a modified Peto's test with

$\rho = 0$, the Peto test with COD information assigned by pathologists, and the Poly-3 test. In this analysis the pairwise test (control versus 30 ppm) and the one-sided trend tests with the lowest 3 groups (control, 30 and 35 ppm) and the lowest 4 groups (0, 30, 35 and 45 ppm) are considered in the analysis. The overall tumor rate on this subset of data is 6.67%.

All tests showed a significant dose-related trend in the four-group trend test ($p < 0.05$, Table 9). Figure 2 displays the survival estimates $\hat{S}(t)$ obtained from the NPMLE method discussed in Section 2.1 using the NTP intervals discussed in Section 3. These functions are obtained from the same maximum likelihood estimators used to impute the numbers of fatal tumors \hat{d}_j for the weight-adjusted Peto test. Although Figure 2 supports a general dose-related trend from 0 to 45 ppm, it does indicate an inversion between the 30 and 35 ppm dose groups. This is reflected in the three-group trend test, in that all p -values are an order of magnitude higher than in the four-group test, even though the trend is not necessarily monotonic in dose. The combined observed tumor rate in the three lowest groups is 5.9%, which is close to the limit where the normal approximation of the Peto test statistic becomes unreliable. In this case, it might be a positive characteristic of the weight-adjusted Peto test that it shows only a marginal effect at the 5% level. For the comparison of 0 and 30 ppm, all four tests gave a result intermediate to the three-group and four-group trend tests, in keeping with the relative positions of the time-to-onset survival distributions in Figure 2.

(Tables 8 and 9 and Figure 2 are here.)

5 Discussion

The proposed methodology is for an animal bioassay that does not have COD information. Our weight-adjusted Peto's test was implemented by estimating tumor lethality via a modification of the NPMLE method of Ahn et al. (2000) when COD was not available. The proposed test was incorporated with a Fleming-Harrington-type weight function via a bootstrap resampling method.

A Monte Carlo simulation was conducted to assess size and power of the proposed test. Results

from the proposed testing method were close to those from Peto's test with small COD assignment error from the simulation studies. We found that the proposed test procedure controls the probability of a Type I error reasonably well by applying the Fleming-Harrington-type weight via the bootstrap resampling method associated with the estimation of tumor lethality given in Moon et al. (2002). In this study, the proposed testing method controlled the probability of a Type I error better than the Poly-3 test for Weibull tumor onset data with a shape parameter of 1.5, and performed as well as that for Weibull distributions with shape parameters of 3.0 and 6.0. Overall, the weight-adjusted Peto test performed better than that without adjustment.

We applied the proposed testing procedure to a subset of the ED₀₁ data to find the low dose effect of feeding 2-AAF to female BALB/C mice. The weight-adjusted Peto test showed a significant dose-related trend ($p < 0.05$) in the development of liver tumors over the dose range from 0 to 45 ppm. However, the apparent inversion in tumor-onset rates between 30 and 35 ppm in Figure 2 caused the three-group trend test to be only marginally significant at the 5% level ($p = 0.0409$). Among the tests compared, the weight-adjusted Peto test showed the least significant for the three-group trend test, which might be interpreted as a positive characteristic given the relative positions of the time-to-onset survival distributions in Figure 2.

Acknowledgements

The work of Hojin Moon and J. Jack Lee was supported in part by the Tobacco Settlement Funds as appropriated by the Texas State Legislature. Hongshik Ahn's work was supported in part by NIH grant 1 R29 CA77289-05.

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.
- 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.
- Fleming, T. R. and Harrington, D. P. (1981). A class of hypothesis tests for one and two samples of censored survival data. *Communications in Statistics*, **10**, 763-794.
- Kodell, R. L. and Ahn, H. (1997). An age-adjusted trend test for the tumor incidence rate. *Biometrics*, **53**, 1467-1474.
- 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.
- 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.
- 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*. In press.
- Moon, H., Ahn, H. and Kodell, R. L. (2002). Extension of Peto's test by attribution of tumor lethality in the absence of cause-of-death information. *Biometrical Journal*. In press.
- 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.

Table 1: Frequency tabulation of number of mice in the ED₀₁ study of 2-AAF

Dose (ppm)	j^a	<u>Natural death</u>		<u>Sacrifice</u>		
		\tilde{d}_j^b	$(a_{1j} + d_j)^c$	b_{1j}^d	a_{2j}^e	b_{2j}^f
0	1	0	0	9	0	0
	2	0	0	15	0	0
	3	1	1	34	0	0
	4	1	1	60	7	137
30	1	0	0	17	0	0
	2	0	2	42	0	0
	3	1	6	67	0	0
	4	2	7	84	22	279
35	1	0	1	9	0	0
	2	3	3	31	0	0
	3	1	1	55	0	0
	4	0	1	80	18	192
45	1	0	0	7	0	0
	2	1	1	13	0	0
	3	3	5	43	0	0
	4	2	3	66	19	132
60	1	0	0	7	0	0
	2	0	2	29	0	0
	3	1	1	46	0	0
	4	1	7	64	28	136
75	1	0	0	10	0	0
	2	0	1	16	0	0
	3	2	3	28	0	0
	4	3	10	41	22	101
100	1	0	0	2	0	0
	2	0	0	14	0	0
	3	1	5	15	0	0
	4	0	3	23	22	51
150	1	0	0	5	0	0
	2	1	2	14	0	0
	3	2	6	24	0	0
	4	7	15	18	16	33

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

^bNumber of fatal tumors assigned by pathologists

^{c,d}Natural death with tumor of interest^c or without tumor^d

^{e,f}Sacrificed but having tumor^e or not having tumor^f

Table 2: Likelihood contribution of each event

Event	Likelihood contribution	Count for interval j
Death from fatal tumor	$Q(t_{j-1})[P(t_{j-1}) - P(t_j)]$	d_j
Death with incidental tumor	$[Q(t_{j-1}) - Q(t_j)][P(t_j) - S(t_j)]$	a_{1j}
Death without tumor	$S(t_j)[Q(t_{j-1}) - Q(t_j)]$	b_{1j}
Sacrifice with tumor	$Q(t_j)[P(t_j) - S(t_j)]$	a_{2j}
Sacrifice without tumor	$S(t_j)Q(t_j)$	b_{2j}

Table 3: Tumor prevalence data for interval I_j .

	Dose group				Total
	1	2	\dots	g	
Number of animals with tumors	y_{1j}	y_{2j}	\dots	y_{gj}	$y_{.j}$
Number of animals without tumors	$n_{1j} - y_{1j}$	$n_{2j} - y_{2j}$	\dots	$n_{gj} - y_{gj}$	$n_{.j} - y_{.j}$
Total deaths from competing risks	n_{1j}	n_{2j}	\dots	n_{gj}	$n_{.j}$

Table 4: Tumor mortality data for interval I_j .

	Dose group				Total
	1	2	\dots	g	
Number of fatal tumor deaths in I_j	x_{1j}	x_{2j}	\dots	x_{gj}	$x_{.j}$
	$m_{1j} - x_{1j}$	$m_{2j} - x_{2j}$	\dots	$m_{gj} - x_{gj}$	$m_{.j} - x_{.j}$
Number of animals surviving in I_j	m_{1j}	m_{2j}	\dots	m_{gj}	$m_{.j}$

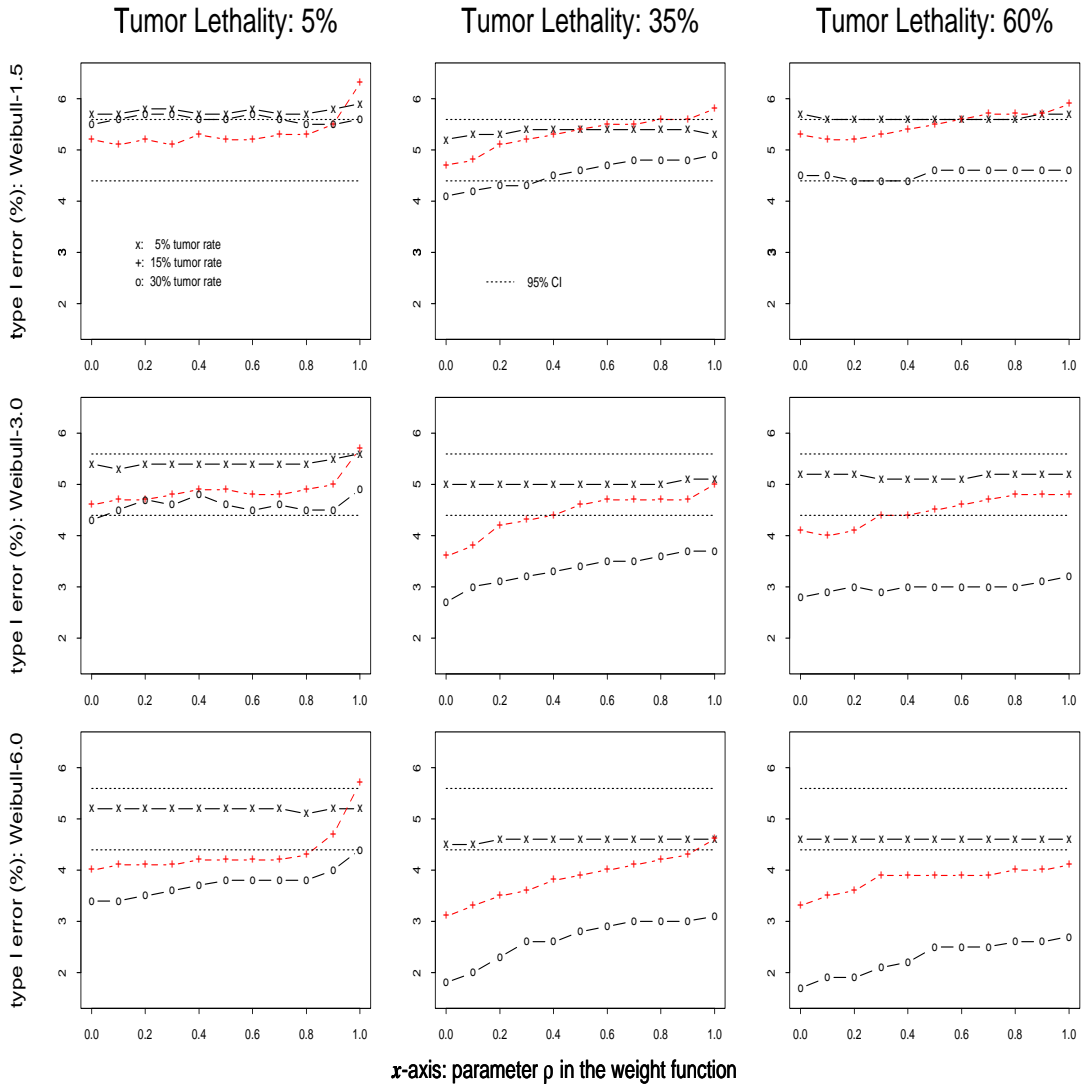


Figure 1: Simulated size of the proposed weight-adjusted Peto test with fixed $\rho \in [0, 1)$. This test reduces to the modified Peto test of Moon et al. (2002) if $\rho = 0$. The average values of ρ^* estimated by our method are given in Tables 5 - 7.

Table 5: Simulated size and power (%) corresponding to nominal 5% significance level for trend tests applied to single-sacrifice data with Weibull tumor onset distribution (with shape parameter 1.5). All tests are based on a dose scaling of 0, 1, 2, and 4 with 5000 trials for each configuration. The competing risks survival rate for each of the 4 groups is 0.6, 0.5, 0.4 and 0.3 for the control, low, intermediate and high dose groups, respectively.

	Tumor Onset Prob. ^a	Tumor Lethal. Prob. ^b	Weight-adjusted Peto test		Poly-3	Peto with correct COD	Peto with Wrong COD		
			$(\rho = 0)^c$	$(\rho^*, \text{s.d.})^d$			30-10 ^e	50-15	70-20
Size	.05	\approx .05	5.7	5.7 (.1128, .1043)	6.1	5.3	7.2	8.1	7.9
	.15	\approx .05	5.2	5.1 (.4408, .0227)	6.2	4.7	6.7	8.8	8.7
	.30	\approx .05	5.5	5.5 (.4152, .0118)	6.9	4.4	8.7	9.6	11.5
	.05	\approx .35	5.2	5.3 (.2873, .0472)	6.1	5.2	6.8	7.3	7.7
	.15	\approx .35	4.7	5.1 (.4248, .0147)	6.2	4.7	6.6	7.8	9.0
	.30	\approx .35	4.1	4.6 (.4176, .0074)	6.9	4.8	7.7	9.9	10.2
							10-30	15-50	20-70
	.05	\approx .60	5.7	5.7 (.2859, .0462)	6.1	5.1	5.7	5.4	4.9
	.15	\approx .60	5.3	5.8 (.4201, .0159)	6.2	4.7	5.0	4.4	3.7
	.30	\approx .60	4.5	4.6 (.4392, .0084)	6.9	5.0	4.5	4.1	3.0
							30-10	50-15	70-20
	.05	\approx .05	90.4	88.6 (.4299, .0214)	91.8	89.1	87.7	88.2	89.9
.15	\approx .05	95.7	93.5 (.4355, .0089)	96.5	94.3	94.1	95.0	95.6	
.30	\approx .05	91.8	89.7 (.4020, .0107)	93.8	89.6	91.2	92.5	95.2	
Power	.05	\approx .35	89.2	86.5 (.4093, .0321)	91.8	89.3	87.6	87.9	88.7
	.15	\approx .35	93.7	91.1 (.4121, .0070)	96.5	94.4	93.5	94.4	95.1
	.30	\approx .35	86.8	88.1 (.4312, .0066)	93.8	88.9	90.0	91.2	92.6
							10-30	15-50	20-70
	.05	\approx .60	89.9	88.8 (.4304, .0338)	91.8	89.7	83.5	80.5	77.4
	.15	\approx .60	94.8	93.7 (.4295, .0074)	96.5	94.7	89.1	85.5	82.2
.30	\approx .60	86.1	88.0 (.4492, .0065)	93.8	89.0	81.7	77.2	73.1	

^aCumulative tumor onset probability at 104 weeks in absence of competing risks

^bTumor lethality rates that actually result in death. Same probability in all dose groups.

^cBefore applying weight function

^dEstimated ρ^* via a bootstrap resampling method and standard deviation

^e $x - y$: $x\%$ chance incidental tumors are classified as fatal;

$y\%$ chance fatal tumors are misclassified as incidental.

Table 6: Simulated size and power (%) corresponding to nominal 5% significance level for trend tests applied to single-sacrifice data with Weibull tumor onset distribution (with shape parameter 3.0). All tests are based on a dose scaling of 0, 1, 2, and 4 with 5000 trials for each configuration. The competing risks survival rate for each of the 4 groups is 0.6, 0.5, 0.4 and 0.3 for the control, low, intermediate and high dose groups, respectively.

	Tumor Onset Prob. ^a	Tumor Lethal. Prob. ^b	Weight-adjusted Peto test		Poly-3	Peto with correct COD	Peto with Wrong COD		
			$(\rho = 0)^c$	$(\rho^*, \text{s.d})^d$			30-10 ^e	50-15	70-20
Size	.05	\approx .05	5.4	5.4 (.2714, .0431)	5.5	5.2	6.8	7.4	7.0
	.15	\approx .05	4.6	4.7 (.4284, .0436)	4.8	4.6	7.0	7.0	7.7
	.30	\approx .05	4.3	4.6 (.4312, .0149)	4.5	3.8	7.0	8.1	9.7
	.05	\approx .35	5.0	5.0 (.2752, .0439)	5.5	5.0	6.7	7.1	6.9
	.15	\approx .35	3.6	4.9 (.4225, .0220)	4.8	4.4	6.4	7.6	7.3
	.30	\approx .35	2.7	3.4 (.4115, .0095)	4.5	4.0	6.6	7.5	8.7
							10-30	15-50	20-70
	.05	\approx .60	5.2	5.2 (.2755, .0430)	5.5	5.1	5.9	6.2	5.2
	.15	\approx .60	4.1	4.9 (.4161, .0102)	4.8	4.6	5.5	4.7	3.9
	.30	\approx .60	2.8	3.0 (.4174, .0144)	4.5	4.3	4.6	3.7	3.2
							30-10	50-15	70-20
	.05	\approx .05	86.3	84.4 (.4016, .0634)	87.5	85.5	84.3	84.2	85.7
.15	\approx .05	92.1	91.0 (.4452, .0122)	93.0	91.7	91.1	91.8	93.2	
.30	\approx .05	87.7	86.0 (.4094, .0132)	88.4	86.8	87.9	89.3	91.7	
Power	.05	\approx .35	84.3	81.5 (.4025, .0525)	87.5	85.9	84.5	84.2	85.0
	.15	\approx .35	88.3	87.8 (.3981, .0124)	93.0	92.2	91.0	91.1	92.0
	.30	\approx .35	78.9	79.8 (.4088, .0083)	88.4	86.4	87.1	88.1	89.5
							10-30	15-50	20-70
	.05	\approx .60	85.6	79.5 (.4165, .0229)	87.5	86.2	79.3	76.6	73.1
	.15	\approx .60	89.3	86.5 (.4191, .0070)	93.0	92.8	86.4	82.0	78.3
	.30	\approx .60	77.5	77.6 (.4277, .0102)	88.4	87.5	78.1	73.1	66.9

^aCumulative tumor onset probability at 104 weeks in absence of competing risks

^bTumor lethality rates that actually result in death. Same probability in all dose groups.

^cBefore applying weight function

^dEstimated ρ^* via a bootstrap resampling method and standard deviation

^e $x - y$: $x\%$ chance incidental tumors are classified as fatal;

$y\%$ chance fatal tumors are misclassified as incidental.

Table 7: Simulated size and power (%) corresponding to nominal 5% significance level for trend tests applied to single-sacrifice data with Weibull tumor onset distribution (with shape parameter 6.0). All tests are based on a dose scaling of 0, 1, 2, and 4 with 5000 trials for each configuration. The competing risks survival rate for each of the 4 groups is 0.6, 0.5, 0.4 and 0.3 for the control, low, intermediate and high dose groups, respectively.

	Tumor Onset Prob. ^a	Tumor Lethal. Prob. ^b	Weight-adjusted Peto test		Poly-3	Peto with correct COD	Peto with Wrong COD		
			$(\rho = 0)^c$	$(\rho^*, \text{s.d.})^d$			30-10 ^e	50-15	70-20
Size	.05	\simeq .05	5.2	5.2 (.0500, .0697)	4.7	5.1	6.4	6.0	6.3
	.15	\simeq .05	4.0	4.1 (.3512, .0785)	3.5	4.0	6.0	6.1	6.2
	.30	\simeq .05	3.4	3.9 (.4328, .0265)	2.4	3.4	5.4	5.7	6.4
	.05	\simeq .35	4.5	4.5 (.0620, .0797)	4.7	4.9	5.9	6.7	5.9
	.15	\simeq .35	3.1	4.6 (.4260, .0213)	3.5	4.2	5.8	6.6	6.3
	.30	\simeq .35	1.8	3.2 (.4195, .0118)	2.4	3.9	5.6	6.5	6.5
Power	.05	\simeq .60	4.6	4.6 (.0653, .0815)	4.7	5.1	6.1	6.3	5.5
	.15	\simeq .60	3.3	4.1 (.4121, .0362)	3.5	4.5	5.5	4.9	3.9
	.30	\simeq .60	1.7	2.5 (.4049, .0109)	2.4	4.2	4.3	4.0	3.0

^aCumulative tumor onset probability at 104 weeks in absence of competing risks

^bTumor lethality rates that actually result in death. Same probability in all dose groups.

^cBefore applying weight function

^dEstimated ρ^* via a bootstrap resampling method and standard deviation

^e $x - y$: $x\%$ chance incidental tumors are classified as fatal;

$y\%$ chance fatal tumors are misclassified as incidental.

Table 8: Frequency data from ED₀₁ study

Dose (ppm)	j^a	\tilde{d}_j^b	\tilde{d}_j^c	Natural death		Sacrifice	
				$(a_{1j} + d_j)^d$	b_{1j}^e	a_{2j}^f	b_{2j}^g
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
30	1	0	0	0	17	0	0
	2	0	0	2	42	0	0
	3	1	1	6	67	0	0
	4	0	2	7	84	22	279
35	1	0	0	1	9	0	0
	2	0	3	3	31	0	0
	3	0	1	1	55	0	0
	4	0	0	1	80	18	192
45	1	0	0	0	7	0	0
	2	0	1	1	13	0	0
	3	1	3	5	43	0	0
	4	0	2	3	66	19	132

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

^bNumber of fatal tumors estimated by our proposed method

^cNumber of fatal tumors assigned by pathologists

^{d,e}Natural death with tumor of interest^d or without tumor^e

^{f,g}Sacrificed but having tumor^f or not having tumor^g

Table 9: Test results for the proposed weight-adjusted Peto test, Peto's test with COD assigned by pathologists and the Poly-3 trend test for the ED₀₁ data given in Table 8. All tests are one-sided.

		Weight-adjusted Peto's test ^a		Peto ^b	Poly-3
Trend test (0, 30, 35, 45) ppm	Z value	2.728 ($\rho = 0$)	2.612 ($\rho^* = .42$)	2.739	2.724
	p -value	.0032	.0045	.0031	.0032
Pairwise test (0, 30) ppm	Z value	2.093 ($\rho = 0$)	2.044 ($\rho^* = .33$)	2.072	2.068
	p -value	.0182	.0205	.0192	.0193
Trend test (0, 30, 35) ppm	Z value	1.864 ($\rho = 0$)	1.741 ($\rho^* = .38$)	1.899	1.864
	p -value	.0312	.0409	.0288	.0312

^aWeight-adjusted Peto's test with estimated tumor lethality via bootstrap

^bPeto's test with pathologist-assigned tumor lethality

ED01 Study: Single Sacrifice Data

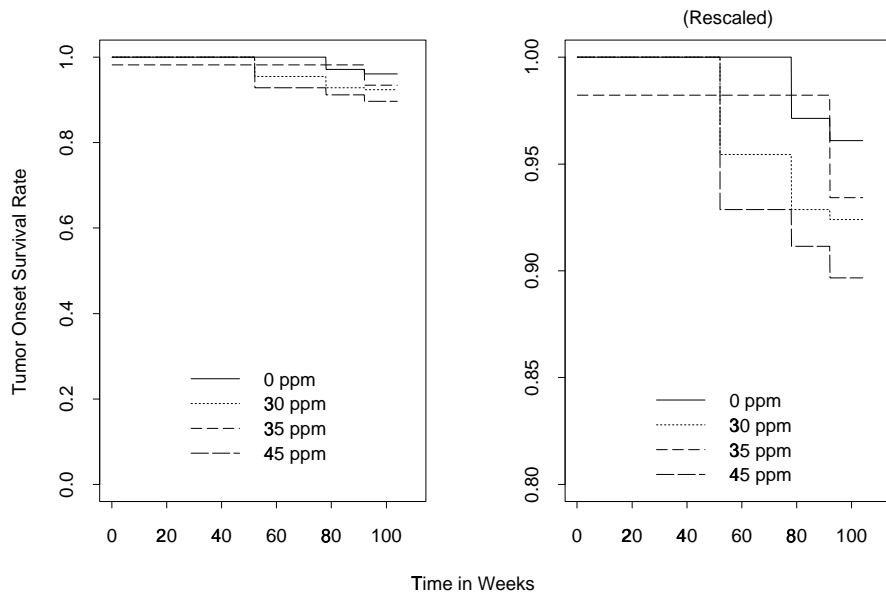


Figure 2: Time-to-tumor-onset survival estimates obtained from the method in Section 2.1 for ED_{01} data. The plot on the right is a rescaled version of the plot on the left.