

Age-Adjusted Exact Trend Tests in the Event of Rare Occurrences

Jessica Y. Mancuso^{1*}, Hongshik Ahn², James J. Chen³, and James P. Mancuso¹

¹Pfizer Global Research and Development, Groton, CT 06340, U.S.A.

²Department of Applied Mathematics and Statistics,
State University of New York at Stony Brook, Stony Brook, NY 11794, U.S.A.

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

* *email*: jessica_y_mancuso@groton.pfizer.com

SUMMARY. Pre-clinical animal carcinogenicity studies are usually concerned with testing the statistical significance of a dose-response relationship. When the response consists of a rare event such as the development of a certain type of tumor, exact statistical methods are often employed. The exact randomization trend test based on the multivariate hypergeometric distribution is less powerful in the presence of treatment-related risks other than the specified response. Particularly, the loss of power becomes more pronounced when competing risks cause progressively higher mortality rates with increasing dose, which is usual in practice. An age-adjusted form of the randomization test is proposed to adjust for this effect. Permutational distribution for Peto's cause-of-death (COD) test is also explored and compared to its asymptotic counterpart by simulation. The use of COD information has been a controversial issue due to the subjectivity in the pathologists' determinations as well as economic reasons. The proposed age-adjusted exact test does not require COD, and it is shown to compare favorably to the COD tests via an extensive Monte Carlo simulation. Applications of the methods to two real data sets are included.

KEY WORDS: Dose-response; Exact randomization test; Exact Poly-3 test; Permutation; Peto cause-of-death test; Survival adjustment.

1 Introduction

Toxicology and risk assessment is an essential component of drug development in the biopharmaceutical industry. Often, animal experiments are conducted to evaluate toxic or carcinogenic risk related to a test compound. In such experiments, animals are randomly allocated to groups including a zero-dose control and two to three groups receiving various dose levels of the compound throughout the course of study. For each animal, the age-at-death and the presence or absence of a particular type of tumor are recorded. All surviving animals are sacrificed and subjected to necropsy at the end of the study. The objective of this study is to test for a dose-related trend in tumor incidence rates at study termination while taking into account the presence of competing risks which are causes of death other than the tumor of interest.

It frequently occurs in animal bioassays that the number of animals developing the tumor of interest is small. There has been some evidence in the literature showing that asymptotic tests may be unreliable in the context of sparse data (Chen and Gaylor, 1986; Ali, 1990). In such situations, exact methods are usually employed. This paper explores the performance of an exact randomization trend test procedure (Ciminera, 1985; Ciminera et al., 1987) that has been routinely used by the drug industry and FDA. The exact size and power for this test under standard experimental settings are computed and tabulated in this paper. In a Monte Carlo simulation, we have found that the empirical performance of the exact randomization trend test is in close agreement with its theoretical size and power if the animals experience no risks other than the tumor of interest. In the presence of competing risks, however, the theoretical size and power overestimate the performance of the test. Particularly, the loss of power becomes larger when the survival patterns due to competing risks differ across dose groups. The primary purpose of this paper is therefore to introduce an age-adjusted

exact method to reduce biases caused by the competing risks. The proposed test can be regarded as an exact version of the Poly-3 test (Bailer and Portier, 1988). The development of this approach was motivated by an analysis of the data from a study of C.I. Acid Red 114 conducted by the NTP (National Toxicology Program, 1991).

The study was conducted by administering desalted, industrial grade C.I. Acid Red 114 in drinking water to groups of female rats for two years. The data set related to the squamous cell carcinoma in the oral cavity epithelium is displayed in Table 1. The sample sizes are (50, 35, 65, 50) and the corresponding dose levels are (0, 150, 300, 600) ppm. All remaining animals were sacrificed during the week of day 729. Exact methods can be applied to test for a dose-related trend on the tumor response since there were only a few tumor-bearing animals. The survival patterns appear to be markedly different across the groups. For instance, only about 50% of the animals in the highest dose group were still alive when the first death occurred in the control and the lowest dose groups. Figure 1 shows the crude and survival-adjusted proportions of tumor-bearing animals versus dose. For a given group, the crude tumor proportion is the rate of tumor cases among all animals in the group. The survival-adjusted tumor proportion results from a time-at-risk weighting adjustment that will be discussed in Section 3.2. One can see the difference between the two curves. By applying the proposed approach to the data, the effect of C.I. Acid Red 114 will be investigated, and the possible existence of any dose-related trend will be examined.

Some other methods have been proposed in analyzing rodent bioassays with respect to a specific disease whose information is confounded with information about the effect of all diseases on mortality. Peto (1974) and Peto et al. (1980) proposed an approach that requires cause-of-death (COD) information for each animal to be determined by pathologists. This method assumes that pathologists can determine whether a tumor has affected an animal's risk of death and correctly classify each tumor as either incidental or fatal. Peto's COD

test is a pooled test statistic resulting from performing tests separately for incidental and fatal tumors. McKnight and Wahrendorf (1992) mentioned in SUMMARY that they relate Peto's test to hypotheses about differences in tumor incidence rate. Janssen (1991) derived exact conditional distributions for various survival tests, including the log-rank test, which is essentially the fatal part of Peto's test. Heimann and Neuhaus (1998) introduced the permutational distribution of the log-rank test statistic for two-group comparisons. In this paper, an approach similar to the permutational distribution for the log-rank test will be applied to Peto's COD test in a general setting of more than two groups.

The use of COD information in analysis of animal bioassays, however, has been controversial because cause-of-death determinations require subjective decisions by pathologists. Misclassification of COD can produce biases for tests (Lagakos, 1982; Lagakos and Louis, 1988). Archer and Ryan (1989) modified Peto's test to account for misclassification of COD. Their test statistic is a function of the misclassification probabilities that may be estimated in certain situations. Dinse (1988) proposed a semiparametric approach to estimate tumor incidence rates when COD is not available, as well as a nonparametric procedure for data with partial information on COD. Malani and Van Ryzin (1988) suggested a nonparametric method to estimate and test the tumor incidence function. Their maximum likelihood estimation method, however, can result in a negative estimate of the tumor incidence rate. This was recognized by Ahn and Kodell (1995), who then introduced a modification for the maximizing procedure.

Our proposed age-adjusted exact method, intended for use with sparse data, will enhance the statistical analysis in toxicology and carcinogenesis studies. With its survival adjustment, the proposed test allows data to be analyzed more accurately by reducing bias in the presence of potential confounders. Moreover, it can be used in general situations regardless of the availability of COD information.

2 The Problem

Consider a single-sacrifice carcinogenicity experiment with g groups of animals. Suppose N animals are initially placed on the experiment, and N_i animals are assigned randomly to group i receiving a dose level d_i of the test substance ($i = 1, 2, \dots, g$). The animals are followed over time for the development of a certain tumor with rare occurrence. Let p_i denote the tumor incidence rate for group i at the study termination. Assuming all animals come from the same population and are born tumor-free on day zero of the experiment, our objective is to test for a dose-related trend in tumor incidence rates at the terminal sacrifice time. The tumor data can be summarized as follows:

Dose Level	d_1	d_2	\dots	d_g	Total
# with tumors	x_1	x_2	\dots	x_g	x
# without tumors	$N_1 - x_1$	$N_2 - x_2$	\dots	$N_g - x_g$	$N - x$
# subjects	N_1	N_2	\dots	N_g	N

Note that the g groups are in order of increasing dose levels $0 = d_1 < d_2 < \dots < d_g$.

3 Exact Randomization Trend Test with a Survival Adjustment

3.1 Exact Randomization Trend Test

The exact randomization trend test, developed by Tukey (Ciminera et al., 1987) and reported by Ciminera (1985), has been routinely employed by the drug industry and FDA for the evaluation of animal carcinogenicity data with few tumor-bearing rodents.

Let random variable X_i indicate the number of positive responses for group i . Conditional on the row margins $X = \sum_i X_i$ and $N = \sum_i N_i$, under the null hypothesis of no treatment effect among the g groups (i.e., $H_0 : p_1 = p_2 = \dots = p_g \equiv p_0$), the distribution

of (X_1, X_2, \dots, X_g) is multivariate hypergeometric. The test is based on the trend score $S = \sum_{i=1}^g d_i X_i$. Denote the value of the trend score and its occurrence probability associated with the observed outcome (x_1, x_2, \dots, x_g) by $s^* = \sum_{i=1}^g d_i x_i$ and $p^* = \left\{ \prod_{i=1}^g \binom{N_i}{x_i} \right\} / \binom{N}{x}$, respectively. The exact one-tailed p -value for a positive trend associated with (x_1, x_2, \dots, x_g) is then

$$P(S \geq s^* \mid x, H_0) = \sum_{\omega_x(s^*)} \frac{\prod_{i=1}^g \binom{N_i}{x_i}}{\binom{N}{x}}, \quad (1)$$

where

$$\omega_x(s^*) = \left\{ (x_1, x_2, \dots, x_g) : 0 \leq x_i \leq N_i, \sum_i x_i = x, \text{ and } \sum_i d_i x_i \geq s^* \right\}. \quad (2)$$

For any given level α , the theoretical size and power for the exact randomization trend test can be computed conditionally (on $x = \sum_i x_i$) or unconditionally. Given α and conditional on x in the observed contingency table, let $t_\alpha(x)$ be the smallest possible cut-off value such that $P(S \geq t_\alpha(x) \mid x, H_0) \leq \alpha$. The value $t_\alpha(x)$, as the conditional critical value of the score at level α , determines the rejection region for the test. The probability $\beta_0(x) = P(S \geq t_\alpha(x) \mid x, H_0)$ is the conditional size of the exact randomization trend test. Similarly, for specified values of p_i 's in the alternative hypothesis H_1 , the conditional power at a given α level is

$$\beta_1(x) = P(S \geq t_\alpha(x) \mid x, H_1) = \sum_{\omega_x(t_\alpha)} \frac{\prod_{i=1}^g \binom{N_i}{x_i} p_i^{x_i} (1 - p_i)^{N_i - x_i}}{\sum_{\Omega_x} \prod_{i=1}^g \binom{N_i}{x_i} p_i^{x_i} (1 - p_i)^{N_i - x_i}},$$

where $\Omega_x = \{(x_1, x_2, \dots, x_g) : 0 \leq x_i \leq N_i \text{ and } \sum_i x_i = x\}$.

The conditional size (power), however, may not be helpful when planning an experiment since the row margin would not be known. Thus, it is useful to compute the unconditional size (power) as a weighted average of the conditional sizes (powers) corresponding to each possible value of the row margin. Define $\pi_0(x) = P(\sum_{i=1}^g X_i = x \mid H_0)$ and $\pi_1(x) = P(\sum_{i=1}^g X_i = x \mid H_1)$ as the probability of observing x under the null and alternative hypotheses, respectively. Note that $\pi_0(x) = \binom{N}{x} p_0^x (1 - p_0)^{N-x}$ and $\pi_1(x) = \sum_{\Omega_x} \prod_{i=1}^g \binom{N_i}{x_i} p_i^{x_i} (1 - p_i)^{N_i - x_i}$. The unconditional size and power of the exact trend test then can be obtained by calculating the

value of the function $\beta(j) = \sum_{x=0}^N \pi_j(x) \beta_j(x)$ with $j = 0$ for size and $j = 1$ for power.

3.2 Proposed Survival Adjustment

Intercurrent mortality refers to all deaths not related to the particular class of tumors which is the focus of study. The causes of intercurrent mortality are referred to as competing risks. If the animals experience no competing risks, then the performance of the exact randomization trend test can be evaluated theoretically. However, the presence of competing risks is usual and therefore the exact test may not achieve its theoretical size and power. Under such circumstances, it is essential to identify and adjust for possible differences in intercurrent mortality among dose groups to eliminate or reduce biases caused by these differences.

Cochran (1954) and Armitage (1955) introduced a well-known test for detecting a linear trend across dose groups in tumor incidence rates at study termination. This test needs an assumption that all animals are at equal risk of developing a tumor over the duration of study. Using the notation introduced earlier, the expected number of animals with tumors in group i is $E_i = N_i x/N$. Defining D_i as $x_i - E_i$, the test statistic for possible monotonic trend with dose is based on $T = \sum_{i=1}^g d_i D_i$, and the variance is estimated by $V = \{x(N-x)/N(N-1)\} \sum_{i=1}^g N_i (d_i - \bar{d})^2$, where $\bar{d} = \sum_{i=1}^g N_i d_i / N$. The Cochran-Armitage (CA) test statistic is $Z_{CA} = T/\sqrt{V}$, where Z_{CA} is asymptotically distributed as standard normal under the null hypothesis of equal tumor incidence rates at study termination among groups. Bailer and Portier (1988) proposed the Poly-3 test, which made an adjustment of the CA test by using a fractional weighting scheme for animals not at full risk of tumor development. Define the number at risk for group i as the sum of N_i weights:

$$r_i = \sum_{j=1}^{N_i} w_{ij}, \quad (3)$$

where w_{ij} is the time-at-risk weight for the j^{th} animal in group i . The risk weight w_{ij} is defined as

$$w_{ij} = \begin{cases} 1 & \text{if the animal dies with the tumor} \\ \left(\frac{t_{ij}}{t_{\max}}\right)^3 & \text{otherwise} \end{cases}$$

where t_{ij} is the animal's death time and t_{\max} is the time to study termination (e.g., terminal sacrifice time). This weighting gives less weight to a tumor-free animal that dies before the end of the study. The test statistic for the Poly-3 test is obtained by replacing N_i with r_i and $N = \sum_i N_i$ with $r = \sum_i r_i$ in the computation of D_i , E_i , V , and Z_{CA} .

We propose a survival adjustment of the exact randomization trend test based on the weighting scheme of the Poly-3 test. Define $[r_i]$ to be the largest integer that does not exceed r_i , where r_i is expressed in Equation (3). The adjusted exact test results from replacing N_i by $[r_i]$ and N by r in Equations (1) and (2).

4 Permutational Distribution for Peto's COD Test

4.1 Peto's COD Trend Test

Peto (1974) and Peto et al. (1980) recommend that pathologists separate observed tumors of interest into two disjoint classes: fatal and incidental. If a tumor is judged to have been responsible for an animal's death, then it is classified as fatal; otherwise, it is classified as incidental. All tumors found in sacrificed animals are classified as incidental. The analysis using cause-of-death information is performed separately for incidental and fatal tumors.

The time span will be divided into m intervals such that the j^{th} interval is $I_j = (t_{j-1}, t_j]$, $j = 1, \dots, m$, where $t_0 = 0$ and t_m denotes the time of the terminal sacrifice. The time points are selected differently for the incidental and fatal parts. First, consider only the animals that died without the tumor or with incidental tumors. Let n_{ij} be the number of these

animals in group i dying during interval I_j , and y_{ij} be the number of these animals that died with incidental tumors. The time span is divided according to the NTP intervals (in weeks) by Bailer and Portier (1988): 0 – 52, 53 – 78, 79 – 92, 93 – 104, and TS corresponding to animals terminally-sacrificed at 104 weeks. The expected number of tumor-bearing animals in group i for interval j is $E_{ij} = y_j K_{ij}$, where $K_{ij} = n_{ij}/n_j$. Thus, the observed and expected numbers of tumor-bearing animals in group i over the entire experiment are $O_i = \sum_{j=1}^m y_{ij}$ and $E_i = \sum_{j=1}^m E_{ij}$, respectively. Define $D_i = O_i - E_i = \sum_{j=1}^m (y_{ij} - E_{ij})$ and $V_{ri} = \sum_{j=1}^m \kappa_j K_{rj} (\delta_{ri} - K_{ij})$, 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 $\mathbf{D}_a = (D_1, \dots, D_g)^\top$ and \mathbf{V}_a be the $g \times g$ matrix with (r, i) entry V_{ri} .

Second, consider the animals with fatal tumors. The method used is similar to that for the incidental tumors, except that each tumor-death time defines an interval. Let m_{ij} be the number of animals in group i surviving at the beginning of the j^{th} interval, and x_{ij} be the number of these animals dying from the tumor in that interval. A vector \mathbf{D}_b of differences of observed and expected values using the above data is calculated the same way as for the incidental tumors, except that x_{ij} is used instead of y_{ij} , and m_{ij} is used instead of n_{ij} . Similarly, the corresponding covariance matrix \mathbf{V}_b can be 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 $Z_{Peto} = \mathbf{l}^\top \mathbf{D} / \sqrt{(\mathbf{l}^\top \mathbf{V} \mathbf{l})}$ can serve as a trend test, where $\mathbf{l} = (d_1, \dots, d_g)^\top$. Under the null hypothesis of no treatment effect among dose groups, Z_{Peto} is asymptotically distributed as standard normal.

4.2 Permutational Distribution

If the number of tumor-bearing animals is large, then the asymptotic version of Peto's test is applicable, but exact methods are preferable for sparse data. Permutation tests consider all possible assignments of animals to dose groups as equally likely, while fixing

the rest of the information obtained in the experiment. Under the null hypothesis of no treatment effect, this results in an exact conditional distribution of the test statistic when intercurrent mortality patterns are equal across groups, and it is asymptotically correct when the mortality patterns are unequal (Fairweather et al., 1998; Heimann and Neuhaus, 1998).

Given a data set, consider all, say M , possible allocations of animals to groups while keeping the observed data for each animal fixed. Corresponding to these M arrangements, we may obtain M values of the test statistic. The permutational distribution of the Peto test statistic results from assigning equal probability to each of these M values. Letting Z_{Peto}^* be the observed value, the p -value is the proportion of the M values that are at least as extreme as Z_{Peto}^* . By exhaustive enumeration, the computation for the p -value using the permutational distribution for the test is straightforward and efficient if the number of animals in the study is small. For data involving large numbers of subjects, the p -value associated with the permutational distribution of the test statistic may be approximated by sampling from the set of all permutations.

5 Simulations and Results

A Monte Carlo simulation study was conducted to evaluate the following tests: the original and adjusted exact randomization tests, Peto's COD test, and the modified Poly-3 test (Bieler and Williams, 1993) which has been adopted by the NTP.

5.1 Design of Monte Carlo Simulation

A typical bioassay design with four groups of 50 animals each and an experimental duration of 104 weeks is used in the study. The design is simulated to have a single terminal sacrifice at the end of the experiment as in the customary long-term rodent bioassay. The dose levels

used are (0, 1, 2, 4) across groups. For Peto's COD test, the NTP intervals are used in the incidental tumor analysis. The intervalization for the fatal tumor analysis is defined by the time of each death from fatal tumors.

The three independent variables T_1 (time to tumor onset), T_2 (time from tumor onset until death from the tumor) and T_3 (time until death from a competing risk) are used to model animal tumorigenicity data. These variables are generated from the modified Weibull distributions used by Portier, Hedges, and Hoel (1986) and others in the literature (Kodell, Chen, and Moore, 1994; Ahn and Kodell, 1995; Kodell and Ahn, 1997; Chang, Ahn, and Chen, 2000). The survival function for T_1 is

$$S(t) = \exp[-\delta_1(t/104)^{\delta_2}] \quad (4)$$

with δ_2 equal to 1.5, 3, or 6, and with δ_1 chosen such that the probability of tumor onset by the end of the study attains the desired rate. Since the study is concerned with rare events, tumor rates (TR) between 0.01 and 0.15 are used. The survival function for T_3 is: $Q(t) = \exp[-\phi(\gamma_1 t + \gamma_2 t^{\gamma_3})]$. With $\gamma_1 = 10^{-4}$, $\gamma_2 = 10^{-16}$ and $\gamma_3 = 7.425531$, the value of $\phi \geq 1$ is chosen such that the competing risks survival rate (CRSR) with respect to all causes of death except for the tumor of interest at 104 weeks is either 0.5 for all groups or (0.5, 0.4, 0.3, 0.2) across groups. The control survival rate chosen represents the one recently observed in the NTP studies for male Fischer 344 rats (Haseman et al., 1998), although it is somewhat below average for B6C3F1 mice and female rats in the NTP feeding studies. For simplicity, the survival function for T_2 has the same form as $Q(t)$ with the same values of γ_1, γ_2 and γ_3 . Tumor lethality (LETH) is defined as the percentage of observed tumors that are cause of death. The parameter ϕ is selected to reflect three levels of tumor lethality: low (5%), intermediate (35%), and high (90%). Note that varying degrees of tumor lethality can affect the results of Peto's COD test while it does not influence the performance of the

Exact, S-Exact, and Poly-3 tests.

For each configuration, 10,000 simulated data sets are generated and tested by various methods at the nominal significance level $\alpha = 5\%$. Due to the fact that the results of Exact, S-Exact, and Poly-3 tests are independent of tumor lethality, T_2 was excluded in simulating data for these tests. For Peto's test, the permutational distribution described in Section 4.2 is used as well as its asymptotic counterpart. In conducting the exact version of Peto's test using the permutational distribution, 5,000 permutation samples are generated from each data set in estimating the p -value.

Since the COD methods assume that pathologists can determine if a tumor has affected an animal's risk of death, such tests should reach their optimum performance when the COD is assigned accurately. In practice, however, pathologists often claim that accurate determinations of COD are impossible. Due to this controversy, we include probabilities for misclassification of COD in the simulation. Let $x\%$ be the probability of misclassifying an incidental tumor as fatal, and $y\%$ be the probability of misclassifying a fatal tumor as incidental. For $\text{LETH} = 5\%$ or 35% , the values for (x, y) chosen are $(30, 10)$ and $(50, 15)$. For $\text{LETH} = 90\%$, the values for (x, y) are selected to be $(10, 30)$ and $(15, 50)$. The reason for choosing the opposite error rates for highly lethal tumors is that for such tumors more animals would die from the tumor of interest than from incidental causes. Furthermore, we include the scenarios of regarding all tumors as fatal and regarding all tumors as incidental since these are common practices in the NTP.

5.2 Results

If the animals experience no risks other than the tumor of interest, the exact randomization test (Exact) and the proposed survival-adjusted test (S-Exact) are equivalent, and their size and power can be computed theoretically. The results of these computations at the 5% level

under a variety of dose-response situations were verified by simulation of 10,000 repetitions per configuration. Strong agreement between the theoretical and empirical performance of the test can be seen in Table 2.

In the presence of competing risks, Table 3 contains results for $\delta_2 = 3$ in modeling T_2 , while Table 4 displays results for $\delta_2 = 1.5$ and 6 which are meant to investigate the possible dependency of the proposed exact method on the choice of δ_2 . The results reported are for models with different CRSR across groups. The results for models with the same CRSR for all groups are not reported since the effectiveness of the survival adjustment for the randomization test is more distinguishable in case of different CRSR. Moreover, since Peto's test associated with the permutational distribution obtains slightly lower size and power than its asymptotic version in most cases, we report only the results of the permutational test for erroneous COD cases. Note that for display purposes, the results for the Exact, S-Exact and Poly-3 tests are included in Tables 3 and 4, although T_2 was not involved in simulating the data for these tests.

Results in Table 2 indicate that the Exact test controls the probability of a Type I error under the nominal level in all cases, as expected. Consistent with the nature of exact methods, the conservatism of this test is expected due to the discrete distribution of the test statistic. However, the Exact test becomes even more conservative when competing risks are involved, as shown in Tables 3 and 4. Our proposed S-Exact test shows the advantage of maintaining reasonable size, and for $\delta_2 = 1.5$, the results are favorable to the S-Exact test as compared to the Poly-3 test. The S-Exact test substantially improves the Exact test in controlling size, whereas the Poly-3 test results in an inflation of size in most cases. For $\delta_2 = 6$, S-Exact shows improvement over Exact, but Poly-3 performs slightly better.

Consider the case that cause-of-death information is accurate, that is, the optimum condition for Peto's test. Although not reported, when the CRSR is the same for all groups,

in general Peto’s asymptotic test is too liberal in rejecting the null hypothesis. Its exact version reduces the Type I error rate, but a slight inflation of size still occurs in some cases. In comparison, the proposed S-Exact test maintains size under the nominal level in all cases. In terms of power, both the S-Exact and Peto tests perform quite well, as shown in Table 3. The proposed exact method without the need of COD is comparable to Peto’s test with accurate COD. From the tables, the impact of the four inaccurate COD scenarios (1, 2, F, I) on Peto’s method is evident. Under scenarios 1 and 2, Peto’s test shows high Type I error rate for low lethality and low Type I error rate for high lethality for $\delta_2 = 1.5$ or 3, although it controls size better in case of $\delta_2 = 6$. Notice that if all tumors are regarded as fatal (F), the size of Peto’s test is markedly out of control for low and intermediate lethalties. Similarly, Peto’s test performs poorly if all tumors are regarded as incidental (I) when in fact the lethality is high.

For $\delta_2 = 3$, in addition to the numerical results included in Table 3, the impact of potential misclassification of COD can be visualized in Figure 2. The curves in the figure represent the simulated size for the four tests: the unadjusted (E) and adjusted (SE) exact randomization tests, the permutational (PP) and asymptotic (PA) versions of Peto’s test. Labeled on the X -axis are accurate COD case (0) and the two inaccurate COD scenarios (1, 2) with increasing error rates as described in Section 5.1. With erroneous COD, Peto’s test fails to control the size. For tumors of low and intermediate lethality, the Type I error rate is generally inflated and the bias becomes more pronounced as the tumor rate and COD error rate increase. In case of high lethality, Peto’s test becomes too conservative and loses power with increasing COD error probabilities and tumor rates.

Comparing with the COD test, the proposed exact method is invariant with respect to the degree of tumor lethality, and it outperforms the original randomization test by the survival adjustment. While not requiring COD to be assigned by pathologists, the proposed

test shows its advantage and robustness. Due to the nature of exact methods, the S-Exact test is somewhat conservative as compared to the Poly-3 test. However, the S-Exact test performs reasonably well while its Type I error rate never exceeds the nominal level.

6 Examples

6.1 C.I. Acid Red 114 Data – COD is unavailable

The proposed exact methods are applied to analyze the C.I. Acid Red 114 data introduced in Section 1. Note that Peto’s test is not applicable due to the lack of COD information in the experiment. In the S-Exact test, the adjusted tumor rates are $(0, 0, 7.9, 15.4)\%$ across doses, quite different from the crude rates $(0, 0, 4.6, 4.0)\%$. As shown in Figure 1, the curve obtained using the age adjustment shows a clear dose-response trend whereas the one with the crude tumor rates does not. The proposed exact trend test results in a p -value of 0.009 showing clear evidence of carcinogenic activity of C.I. Acid Red 114, while the original test, with a p -value of 0.099, fails to detect a significant progressive trend.

6.2 ED₀₁ Data – COD is available

The ED₀₁ study, a large experiment intended to investigate the effect of the 2-acetylaminofluorene (2-AAF), was initiated by the National Center for Toxicological Research in 1976 (Littlefield et al., 1980). The study was conducted by feeding 2-AAF to female BALB/C mice. Consider the 24-month cohort consisting of mice that were designated to be sacrificed at 24 months. Due to the large number of mice from this cohort and the concern of the feasibility of exact methods, a data set is obtained by randomly drawing 50 mice from each of the four groups that were dosed continuously at concentrations of 0, 30, 60, 75 ppm 2-AAF. The resulting sample can be regarded as a random sample from the target population.

The associated tumors of interest are hepatocellular adenomas or carcinomas of the liver.

In this experiment, the cause-of-death information was assessed. The counts for these COD assignments are displayed in Table 5. The NTP intervals described in Section 5.1 are used in the analysis for deaths with incidental tumors. The intervalization for the fatal component is defined by time of each death from fatal tumors. Utilizing the COD, Peto's test gives p -values of 0.032 and 0.033 associated with the permutational and asymptotic distributions, respectively. The p -value for the permutational test is based on 10,000 randomly generated permutation samples.

Without the use of COD, the age-adjusted exact randomization test results in a p -value of 0.037 and suggests rejecting the null hypothesis of equal tumor probabilities at $\alpha = 5\%$, while the original randomization test fails to reject the null hypothesis (p -value = 0.060). In agreement with Peto's COD test, the proposed test detects evidence of effect of 2-AAF on development of hepatocellular adenomas or carcinomas of the liver whereas the unadjusted exact randomization test shows no evidence of trend at 5% significance level.

7 Concluding Remarks

A problem for the statistical analysis of tumor incidence rates arises from the presence of treatment-induced mortality unrelated to the tumor of interest. The exact randomization trend test that is commonly employed in the event of rare occurrences could be biased since it focuses on crude lifetime tumor incidence rates and makes no adjustment for differences in survival experiences across dose groups. The age-adjusted form of the test developed in this work can result in great improvement for the analysis of sparse data.

Occasionally, cause-of-death information is assigned in animal tumorigenicity experiments. The use of COD, however, has been a major issue because of its subjectivity. Indeed,

many pathologists are uncomfortable in assigning cause of each death due to the complexity of isolating specific tumorigenic lesions as the COD. Moreover, they are often reluctant to strictly classify each tumor as either incidental or fatal. The reliability of COD information has been questioned by many researchers in the literature. Hence, one should be alert and account for the potential misclassification of COD when evaluating tests utilizing COD. Our results indicate that Peto's test performs quite well if the COD assignments are accurate. Compared with the optimal performance of Peto's test, the proposed exact method demonstrates its merit and robustness by the elimination of the COD requirement.

The Poly-3 test, a currently popular procedure, was included in this work for comparisons. Our results demonstrate that this asymptotic test works fairly well under many circumstances, although its size is inflated for certain cases. The proposed randomization method performs reasonably well while maintaining test size at the nominal level in all cases. The proposed procedure, as an exact version of the Poly-3 test, is highly recommended for use in the event of rare occurrences. The program suitable for data analysis by the proposed method will be provided on the authors' website.

ACKNOWLEDGEMENTS

Jessica Mancuso and Hongshik Ahn's work was supported by National Institute of Health grant R29 CA77289-04. The authors are grateful to the Associate Editor, and two anonymous referees for their helpful comments which greatly improved this paper.

REFERENCES

Ahn, H. and Kodell, R. L. (1995). Estimation and testing of tumor incidence rates in experiments lacking cause-of-death data. *Biometrical Journal* **37**, 745-765.

- Ali, M. W. (1990). Exact versus asymptotic tests of trend of tumor prevalence in tumorigenicity experiments: a comparison of p-values for small frequency of tumors. *Drug Information Journal* **24**, 727-737.
- Archer, L. E., and Ryan, L. M. (1989). Accounting for misclassification in the cause-of-death test for carcinogenicity. *Journal of the American Statistical Association* **84**, 787-791.
- 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.
- Chang, J. Y., Ahn, H., and Chen, J. J. (2000). On sequential closed testing procedures for a comparison of dose groups with a control. *Communications in Statistics, Theory and Methods* **29**(5&6), 941-956.
- Chen, J. J. and Gaylor, D. W. (1986). The upper percentiles of the distribution of the log-rank statistics for small number of tumors. *Communications in Statistics, Simulations* **15**(4), 991-1002.
- Ciminera, J. L. (1985). Some issues in the design, evaluation, and interpretation of tumorigenicity studies in animals. In: *Proceedings of the Symposium on Long-Term Animal Carcinogenicity Studies: A Statistical Perspective*. American Statistical Association 26-35.
- Ciminera, J. L., Heyse, J. F., Oppenheimer, L., and Tukey, J. W. (1987). Randomization methods for tumor sites with few tumor-bearing rodents. *Merck Sharp and*

Dohme Research Laboratories, West Point, PA 19486.

- Cochran, W. G. (1954). Some methods for strengthening the common χ^2 tests. *Biometrics* **10**, 417-451.
- Dinse, G. E. (1988). Estimating tumor incidence rates in animal carcinogenicity experiments. *Biometrics* **44**, 405-415.
- Fairweather, W. R., Bhattacharyya, A., Ceuppens, P. P., Heimann, G., Hothorn L. A., Kodell, R. L., Lin, K. K., Mager, H., Middleton, B. J., Slob, W., Soper, K. A., Stallard, N., Ventre, J., and Wright, J. (1998). Biostatistical methodology in carcinogenicity studies. *Drug Information Journal* **32**, 401-421.
- Haseman, J. K., Hailey, J. R., and Morris, R. W. (1998). Spontaneous neoplasm incidences in Fischer 344 rats and B6C3F1 mice in two-year carcinogenicity studies: A National Toxicology Program Update. *Toxicology Pathology* **26**, 428-441.
- Heimann, G. and Neuhaus, G. (1998). Permutational distribution of the log-rank statistic under random censorship with applications to carcinogenicity assays. *Biometrics* **54**, 168-184.
- Janssen, A. (1991). Conditional rank tests for randomly censored data. *Annals of Statistics* **19**, 1434-1456.
- Kodell, R. L. and Ahn, H. (1997). An age-adjusted trend test for the tumor incidence rate for multiple-sacrifice experiments. *Biometrics* **53**, 1467-1474.
- 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**(4), 959-980.
- 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.
- 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 American Statistical Association* **83**, 1171-1177.
- McKnight, B. and Wahrendorf, J. (1992). Tumour incidence rate alternatives and the cause-of-death test for carcinogenicity. *Biometrika* **79**, 131-138.
- National Toxicology Program (1991). Toxicology and carcinogenesis studies of C.I. acid red 114 in F344/N rats (drinking water studies). *NTP Technical Report 405*, Research Triangle Park, NC.
- Peto, R. (1974). Guidelines on the the analysis of tumor 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, 311-346. International Agency for Research on Cancer: Lyon, France.
- Portier, C., Hedges, J. and Hoel, D. G. (1986). Age-specific models of mortality and tumor onset for historical control animals in the National Toxicology Program's carcinogenicity experiments. *Cancer Research* **46**, 4372-4378.

C.I. Acid Red 114

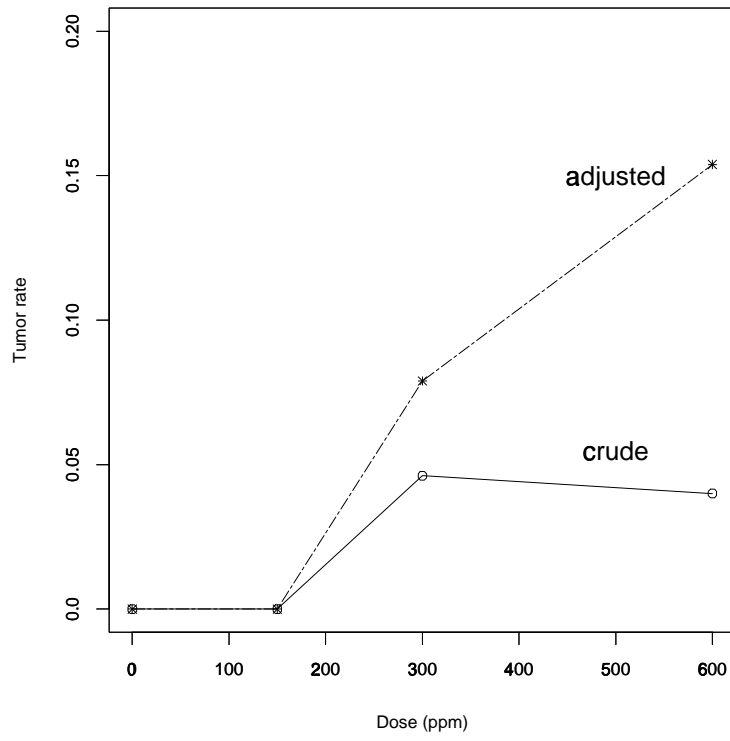


Table 1: *Data on the incidence of squamous cell carcinoma in oral cavity epithelium for female rats in the 2-year drinking water study of C.I. Acid Red 114*

Dose	Survival Time (in days)	
	without tumors at death (frequency in parentheses)	with tumors at death
Control	431, 505, 524, 533, 547, 562, 587, 589, 641, 663, 683, 711, 715(2), 733(5), 734(31)	
150	411(2), 462, 500, 523, 538, 567, 582, 607, 614, 638(2), 641, 658(2), 662, 663, 665, 677, 708, 715, 726, 733(5), 734(8)	
300	53, 87, 285, 344, 400, 403, 448, 477, 487, 491, 494, 496, 515, 517, 521, 523, 531, 532, 538, 545(2), 554, 567, 579, 581, 582, 587, 588, 592(4), 599, 609, 614, 617, 631, 637, 638, 641(2), 648, 649, 656, 658, 676, 677, 680, 693(2), 694, 709, 715(4), 733(5), 734	454, 530, 575
600	186, 229, 258, 276, 339, 344, 349, 351, 355(2), 379, 383, 384(2), 392, 400, 403(2), 411(3), 412, 413(3), 421, 433(3), 454, 456, 458, 473, 475, 477, 486, 488, 501(3), 545, 551(2), 578, 582, 599, 603, 614	372, 411

All deaths at day 729 or later were terminal sacrifices

Table 2: *The theoretical and simulated size and power (%) for the unconditional exact randomization test in the absence of competing risks at $\alpha = 5\%$ significance level for a four-dose experiment with 50 animals per group and doses (0, 1, 2, 4).*

	Tumor rates ^a	Theoretical	Simulated ^b
Size	(.01 .01 .01 .01)	0.78	0.74
	(.02 .02 .02 .02)	2.35	2.33
	(.03 .03 .03 .03)	3.28	3.29
	(.04 .04 .04 .04)	3.80	3.80
	(.05 .05 .05 .05)	4.15	4.14
	(.075 .075 .075 .075)	4.34	4.38
	(.10 .10 .10 .10)	4.41	4.40
Power	(.01 .02 .06 .12)	82.45	82.91
	(.02 .02 .14 .14)	83.51	83.89
	(.01 .03 .14 .12)	77.20	77.51
	(.01 .01 .01 .14)	93.08	92.87
	(.01 .12 .10 .14)	56.86	56.79
	(.01 .14 .14 .12)	36.05	36.33

^a Probabilities across groups of developing the tumor by study termination (104 weeks)

^b Based on 10,000 trials for each configuration

Table 3: *Simulated size and power (%) at $\alpha = 5\%$ significance level based on 10,000 trials for survival-adjusted exact randomization test, Peto's combined trend test with accurate or inaccurate COD, in a four-dose experiment with 50 animals per group and doses (0, 1, 2, 4). The competing risks survival rates are (0.5, 0.4, 0.3, 0.2) across dose groups. $\delta_2 = 3.0$.*

		LETH ^a	Tumor rates	Exact	S-Exact	Poly-3	Peto's Combined Trend Test					
							Accurate COD		Inaccurate COD ^d			
						Perm ^b (Asym ^c)	1	2	F	I		
Size	.05	(.01 .01 .01 .01)		0.23	1.29	2.03	6.38 (6.14)	5.27	4.28	2.13	6.28	
		(.03 .03 .03 .03)		1.81	3.42	4.88	5.06 (5.67)	4.59	5.15	6.03	5.17	
		(.05 .05 .05 .05)		2.01	3.89	5.47	4.25 (4.74)	4.90	5.26	8.12	4.22	
		(.075 .075 .075 .075)		2.48	4.11	5.18	4.70 (4.98)	5.32	6.46	10.06	4.00	
		(.10 .10 .10 .10)		2.10	4.20	4.93	4.11 (4.45)	5.74	6.95	12.26	4.15	
	.35	(.01 .01 .01 .01)		0.23	1.29	2.03	4.83 (4.69)	3.70	3.22	1.90	5.34	
		(.03 .03 .03 .03)		1.81	3.42	4.88	4.47 (5.01)	4.29	4.48	5.31	3.76	
		(.05 .05 .05 .05)		2.01	3.89	5.47	4.52 (4.97)	4.36	4.74	6.82	3.07	
		(.075 .075 .075 .075)		2.48	4.11	5.18	4.16 (4.51)	4.96	5.56	7.84	2.79	
		(.10 .10 .10 .10)		2.10	4.20	4.93	4.19 (4.51)	5.07	5.68	9.67	2.39	
	.90	(.01 .01 .01 .01)		0.23	1.29	2.03	2.33 (2.45)	2.27	2.25	2.09	2.51	
		(.03 .03 .03 .03)		1.81	3.42	4.88	4.64 (5.05)	3.08	2.53	4.65	2.00	
		(.05 .05 .05 .05)		2.01	3.89	5.47	4.05 (4.53)	3.19	2.55	4.87	1.41	
		(.075 .075 .075 .075)		2.48	4.11	5.18	4.19 (4.61)	2.68	2.00	4.80	1.01	
		(.10 .10 .10 .10)		2.10	4.20	4.93	4.54 (4.95)	3.06	1.90	5.16	0.86	
Power	.05	(.01 .02 .06 .12)		58.8	68.3	74.29	70.2 (72.0)	69.7	73.7	78.6	69.9	
		(.02 .02 .14 .14)		59.9	69.6	77.62	69.5 (70.9)	73.7	75.5	83.7	70.1	
		(.01 .03 .14 .12)		52.2	63.6	72.89	64.2 (65.7)	67.2	70.0	79.6	62.6	
		(.01 .01 .01 .14)		74.4	81.2	84.98	84.1 (85.5)	83.7	85.2	86.9	84.5	
		(.01 .12 .10 .14)		31.4	43.3	53.23	46.0 (47.4)	49.9	54.2	66.8	44.4	
	.35	(.01 .02 .06 .12)		58.8	68.3	74.29	69.8 (71.5)	70.0	71.2	76.0	62.7	
		(.02 .02 .14 .14)		59.9	69.6	77.62	70.3 (71.9)	72.4	74.2	81.2	61.1	
		(.01 .03 .14 .12)		52.2	63.6	72.89	62.7 (64.5)	64.9	66.8	75.7	54.8	
		(.01 .01 .01 .14)		74.4	81.2	84.98	82.9 (84.6)	83.1	83.5	86.5	80.1	
		(.01 .12 .10 .14)		31.4	43.3	53.23	44.9 (46.4)	47.4	50.2	60.8	34.5	
	.90	(.01 .02 .06 .12)		58.8	68.3	74.29	69.6 (71.0)	63.5	57.6	70.5	46.4	
		(.02 .02 .14 .14)		59.9	69.6	77.62	71.1 (72.9)	61.8	44.2	72.4	46.1	
		(.01 .03 .14 .12)		52.2	63.6	72.89	64.6 (66.6)	54.9	48.6	65.7	33.2	
		(.01 .01 .01 .14)		74.4	81.2	84.98	84.1 (85.2)	80.0	76.7	85.4	70.6	
		(.01 .12 .10 .14)		31.4	43.3	53.23	45.9 (47.5)	35.9	29.3	48.0	17.2	

^a LETH: proportion of observed tumors that actually result in death. Same probability in all groups

^b Perm: results using permutational distribution based on 5,000 permutation samples

^c Asym: results using asymptotic distribution of the test statistic

^d ($x\%$, $y\%$) = (% of misclassifying incidental tumor as fatal, % misclassifying fatal tumor as incidental)

For LETH = 0.05 or 0.35: (x, y) = (30, 10) and (50, 15) for 1 and 2, respectively.

For LETH = 0.90: (x, y) = (10, 30) and (15, 50) for 1 and 2, respectively.

F: regard all tumors as fatal, I: regard all tumors as incidental

Table 4: *Simulated size (%) at $\alpha = 5\%$ significance level based on 10,000 trials for survival-adjusted exact randomization test, Peto's combined trend test with accurate or inaccurate COD, in a four-dose experiment with 50 animals per group and doses (0, 1, 2, 4). The competing risks survival rates are (0.5, 0.4, 0.3, 0.2) across dose groups. $\delta_2 = 1.5$ or 6.0.*

δ_2^a	LETH ^b	Tumor rates	Exact	S-Exact	Poly-3	Peto's Combined Trend Test					
						Accurate COD		Inaccurate COD ^e			
						Perm ^c (Asym ^d)	1	2	F	I	
1.5	.05	(.01 .01 .01 .01)	0.51	1.89	2.83	5.85 (5.66)	4.43	4.21	2.66	6.07	
		(.03 .03 .03 .03)	2.02	3.72	5.49	5.06 (5.57)	4.95	5.44	7.22	5.00	
		(.05 .05 .05 .05)	2.72	4.72	6.25	4.90 (5.22)	5.47	6.21	9.44	4.13	
		(.075 .075 .075 .075)	2.73	4.92	6.50	4.67 (5.07)	5.50	6.36	11.85	4.08	
		(.10 .10 .10 .10)	2.58	5.00	6.30	4.62 (4.82)	6.41	7.73	14.26	3.88	
	.35	(.01 .01 .01 .01)	0.51	1.89	2.83	4.15 (4.20)	3.52	3.26	2.54	4.35	
		(.03 .03 .03 .03)	2.02	3.72	5.49	4.30 (4.85)	4.49	4.77	5.76	3.07	
		(.05 .05 .05 .05)	2.72	4.72	6.25	4.72 (5.21)	4.04	5.02	7.12	2.48	
		(.075 .075 .075 .075)	2.73	4.92	6.50	4.43 (4.71)	4.85	4.81	7.61	1.96	
		(.10 .10 .10 .10)	2.58	5.00	6.30	4.30 (4.62)	4.91	5.21	8.49	1.84	
	.90	(.01 .01 .01 .01)	0.51	1.89	2.83	2.59 (2.85)	2.27	1.90	2.37	1.70	
		(.03 .03 .03 .03)	2.02	3.72	5.49	4.44 (4.81)	2.97	2.63	4.59	1.66	
		(.05 .05 .05 .05)	2.72	4.72	6.25	4.50 (4.88)	2.97	2.14	4.96	1.17	
		(.075 .075 .075 .075)	2.73	4.92	6.50	4.44 (4.75)	2.76	1.91	4.64	0.90	
		(.10 .10 .10 .10)	2.58	5.00	6.30	4.98 (5.37)	2.39	1.87	4.84	0.53	
6.0	.05	(.01 .01 .01 .01)	0.17	0.85	1.28	6.38 (6.08)	5.06	3.71	1.27	6.62	
		(.03 .03 .03 .03)	0.94	2.28	3.49	5.39 (5.74)	4.80	4.59	4.86	5.17	
		(.05 .05 .05 .05)	1.83	3.17	4.40	4.12 (4.62)	4.31	5.33	6.44	4.77	
		(.075 .075 .075 .075)	1.33	2.56	3.60	4.18 (4.58)	4.40	4.76	8.39	4.27	
		(.10 .10 .10 .10)	1.46	2.65	3.43	3.98 (4.31)	4.52	5.91	9.40	3.84	
	.35	(.01 .01 .01 .01)	0.17	0.85	1.28	5.71 (5.42)	4.16	3.43	1.39	6.16	
		(.03 .03 .03 .03)	0.94	2.28	3.49	4.72 (5.11)	4.22	4.15	4.58	4.54	
		(.05 .05 .05 .05)	1.83	3.17	4.40	4.24 (4.70)	4.14	4.89	6.21	3.95	
		(.075 .075 .075 .075)	1.33	2.56	3.60	3.93 (4.31)	4.39	4.68	7.33	3.54	
		(.10 .10 .10 .10)	1.46	2.65	3.43	3.74 (4.05)	4.33	5.45	8.58	3.13	
	.90	(.01 .01 .01 .01)	0.17	0.85	1.28	2.20 (2.19)	2.50	2.81	1.30	3.58	
		(.03 .03 .03 .03)	0.94	2.28	3.49	3.50 (3.94)	2.93	2.93	4.07	2.76	
		(.05 .05 .05 .05)	1.83	3.17	4.40	3.73 (4.11)	3.15	2.92	4.58	1.92	
		(.075 .075 .075 .075)	1.33	2.56	3.60	4.40 (4.74)	3.02	2.39	4.95	1.68	
		(.10 .10 .10 .10)	1.46	2.65	3.43	4.22 (4.67)	3.20	2.38	5.17	1.55	

^a δ_2 : parameter in Equations (4)

^b LETH: proportion of observed tumors that actually result in death. Same probability in all groups

^c Perm: results using permutational distribution based on 5,000 permutation samples

^d Asym: results using asymptotic distribution of the test statistic

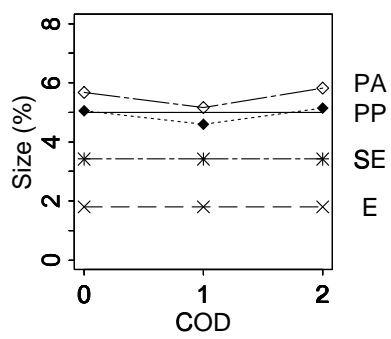
^e $(x\%, y\%) = (\%$ of misclassifying incidental tumor as fatal, $\%$ misclassifying fatal tumor as incidental)

For LETH = 0.05 or 0.35: $(x, y) = (30, 10)$ and $(50, 15)$ for 1 and 2, respectively.

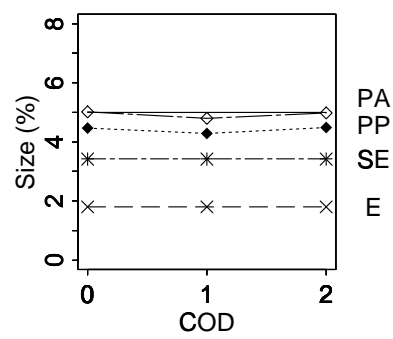
For LETH = 0.90: $(x, y) = (10, 30)$ and $(15, 50)$ for 1 and 2, respectively.

F: regard all tumors as fatal, I: regard all tumors as incidental

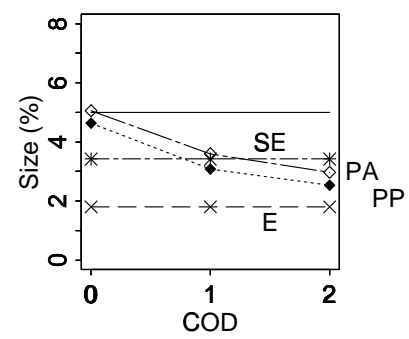
TR = 3%, LETH = 5%



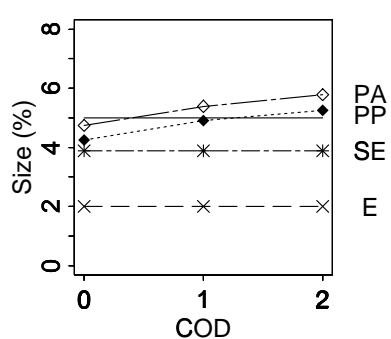
TR = 3%, LETH = 35%



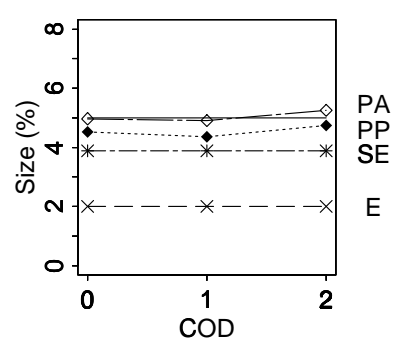
TR = 3%, LETH = 90%



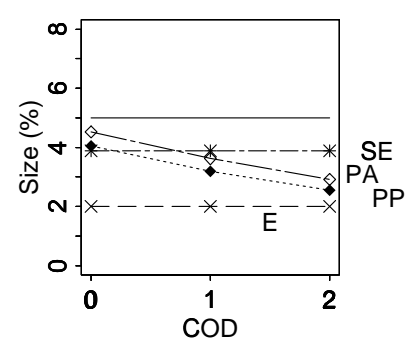
TR = 5%, LETH = 5%



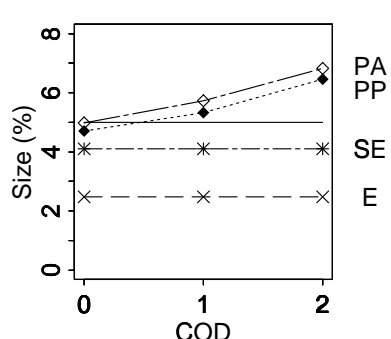
TR = 5%, LETH = 35%



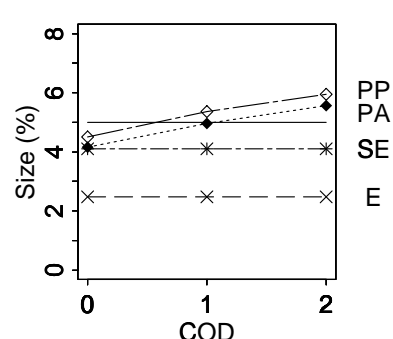
TR = 5%, LETH = 90%



TR = 7.5%, LETH = 5%



TR = 7.5%, LETH = 35%



TR = 7.5%, LETH = 90%

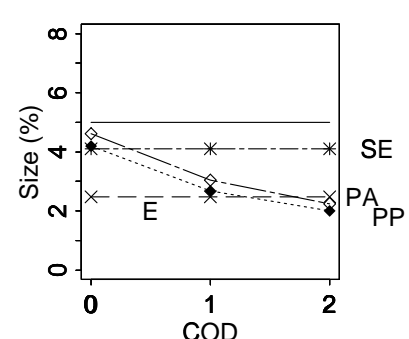


Table 5: *Tumor counts for ED₀₁ data with cause-of-death information*

Tumor	Interval	Time (in days)	Dose (ppm)				Total
			0	30	60	75	
Incidental ^a	1	(0, 364]	0	0	0	0	0
	2	(364, 546]	0	0	0	0	0
	3	(546, 644]	0	1	0	0	1
	4	(644, 729)	2	2	0	0	4
		Sac ^b		0	0	3	5
Fatal	1	(0, 552]	0	0	1	0	1
	2	(552, 623]	0	0	0	1	1
	3	(623, 655]	1	0	0	0	1
	4	(655, 704]	0	0	0	1	1
	5	(704, 719]	0	0	0	1	1
	6	(719, Sac ^b]	0	0	0	0	0
Total			3	3	4	8	18

^a NTP intervals with time points 52, 78, 92, and 104 weeks are used

^b Terminal sacrifice time (all deaths at day 729 or later were terminal sacrifices)

Figure 1: Crude and survival-adjusted tumor proportions in the C.I. Acid Red 114 data for squamous cell carcinoma in the oral cavity epithelium.

Figure 2: Simulated Size (%) at $\alpha = 5\%$ significance level for the exact randomization test (E), survival-adjusted exact test (SE), Peto's cause-of-death (COD) test using permutational distribution (PP) and its asymptotic counterpart (PA) under three COD scenarios: Accurate (0), Inaccurate (1, 2). For low and intermediate tumor lethalties (LETH = 5%, 35%), $(x, y) = (30, 10)$ and $(50, 15)$ for COD = 1 and 2, respectively, where $(x\%, y\%) = (\%$ of misclassifying incidental tumor as fatal, $\%$ of misclassifying fatal tumor as incidental). For high tumor lethality (LETH = 90%), $(x, y) = (10, 30)$ and $(15, 50)$ for COD = 1 and 2, respectively. TR is the probability of developing the tumor by the end of study (104 weeks) for all groups. $\delta_2 = 3$ in survival function of T_1 . The competing risks survival rates are (0.5, 0.4, 0.3, 0.2) across dose groups.