

# ORDER-RESTRICTED DOSE-RELATED TREND TESTS

JESSICA Y. CHANG<sup>1</sup>, HONGSHIK AHN<sup>1\*</sup> AND JAMES J. CHEN<sup>2</sup>

<sup>1</sup>*Department of Applied Mathematics and Statistics, State University of New York at Stony Brook,  
Stony Brook, NY 11794-3600, U.S.A.*

<sup>2</sup>*Division of Biometry and Risk Assessment, National Center for Toxicological Research,  
Food and Drug Administration, Jefferson, AR 72079, U.S.A.*

## SUMMARY

Methods of isotonic regression are applied to increase the power of common trend tests in situations where a monotonicity constraint is imposed upon the dose-response function. Isotonic versions of Cochran-Armitage type trend tests for binary response data are developed, and the bootstrap method is used in finding the empirical distributions of the test statistics and their critical values. The isotonic likelihood ratio test with a survival adjustment is also proposed. This survival adjustment can be applied to the likelihood ratio test for either the order-restricted or unrestricted parameter cases. To achieve the isotonic modifications, an amalgamation algorithm is applied when the observed dose-response is non-monotonic. A Monte Carlo simulation study comparing these trend tests shows the advantages of the isotonic modifications and survival adjustment. By applying the proposed methods to data from a toxicology and carcinogenesis study conducted as part of the National Toxicology Program, the effect of C.I. Pigment Red 23 is investigated.

\*Correspondence to: Hongshik Ahn, Department of Applied Mathematics and Statistics, State University of New York at Stony Brook, Stony Brook, NY 11794-3600, U.S.A. E-mail: hahn@ams.sunysb.edu. Phone: (631)632-8372. Fax: (631)632-8490

Contract/grant sponsor: National Institute of Health  
Contract/grant number: R29 CA77289-03

## 1. INTRODUCTION

Dose-Response experiments are conducted to evaluate toxic or therapeutic treatment effects of a test compound. In such experiments, subjects are randomly allocated to groups receiving various dose levels, and an objective is to determine if any dose-related trend exists in the response variable of interest. Since higher doses are frequently expected to produce treatment effects that are at least as strong as those produced by lower doses, it is common to make the prior assumption that the dose-response curve is monotone.

One primary purpose of this paper is to propose an isotonic regression approach for common trend tests. This approach enables trend tests to be applied under the monotonicity constraint of the dose-response, and is expected to substantially increase the statistical power of the tests while controlling the probability of a Type I error. The development of this approach was motivated by an analysis of the data of C.I. Pigment Red 23 from a toxicology and carcinogenesis study conducted as part of the National Toxicology Program.

Groups of 50 female Fischer 344 rats were fed diets containing 0, 10,000, 25,000, or 50,000 ppm C.I. Pigment Red 23 for two years. All the remaining live animals were sacrificed at the end of the study, and the two tumor types of primary interest were C-cell carcinoma (TGC) and follicular cell adenoma (TGF) of the thyroid gland. The data given in Table I contain individual animal tumor pathology for this study including the survival time and the presence or absence of the tumor at time of death. 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. Notice the non-monotonicity in these curves. Even when the true dose-response relationship is monotone, it is not unusual to observe such non-monotonic sample proportions. In this case, the tests for detecting a linear trend across the dose groups can be weakened by the non-monotonicity in the sample proportions. By applying the proposed approach to these data, the effect of C.I. Pigment Red 23 will be investigated, and the possible existence of any dose-related trend under the order restriction will be examined.

In this paper, we focus on animal carcinogenicity studies where the response of interest for each individual is the presence or absence of a particular type of tumor at death. In most of these experiments, animals that live to the end of the study are sacrificed, but interim sacrifices are not employed. All animals are examined for the existence of the tumor at death, but cause-of-death (COD) information is not collected. The objective is to test for trend in the proportions of animals with the tumor at death in the presence of competing risks. Cochran<sup>1</sup> and Armitage<sup>2</sup> developed a well-known trend test for binary response data. Bailer and Portier<sup>3</sup> proposed an adjustment to the Cochran-Armitage (CA) test to correct for treatment lethality unrelated to the tumor of interest. Their test is called the Poly-3 test, and an improved version of it was given by Bieler and Williams.<sup>4</sup> In Section 3, we propose isotonic versions of the original CA test and the Poly-3 test of Bieler and Williams that can be applied under the monotonicity constraint. The isotonic tests are based on an iterative pairwise averaging process that corrects for non-monotonicity in the observed proportions of tumor-bearing animals. The bootstrap method is used to obtain critical values of these tests.

A likelihood ratio test (LRT) can also be used to test for trend in the proportions, and an isotonic LRT is given by Robertson, Wright, and Dykstra.<sup>5</sup> In this paper, the performance of the standard and isotonic LRTs is explored, and these tests are found to be highly sensitive under varying degrees of treatment lethality independent of the tumor of interest. Another purpose of this paper is to introduce a survival adjustment for these tests. This survival adjustment can be applied to the LRT for either the order-restricted or unrestricted parameter cases. In Section 4, we propose a survival-adjusted LRT and its isotonic counterpart. These tests are also applied to analyze the C.I. Pigment Red 23 data.

A Monte Carlo simulation study was conducted to make comparisons among the CA-type and LRT tests and to determine the benefits of survival adjustments and isotonic modifications. The results are given in Section 5 as well as the details of the simulation design and the bootstrapping method employed. Our proposed methods will enhance the statistical analysis and interpretation of dose-response experiments in toxicology and carcinogenesis studies. The isotonic modification enables researchers to conduct trend tests when a prior monotonicity constraint is imposed upon the dose-response relationship. Not only does this

approach increase the power of common trend tests, but also it facilitates the interpretation of non-monotonic sample dose-response curves if the assumption of monotonicity is correct.

## 2. THE PROBLEM

Consider an animal carcinogenicity experiment with  $g$  dose groups. Let  $d_i$  be the dose level of the  $i^{\text{th}}$  group. Suppose  $n$  animals are initially placed on experiment, and  $n_i$  animals are assigned randomly to the  $i^{\text{th}}$  dose group. The  $n_i$  animals in the  $i^{\text{th}}$  dose group are followed over time for the development of the tumor of interest. We assume that all animals come from the same population and are born tumor-free on day zero of the experiment. With a terminal sacrifice at the end of study, the tumor data can be summarized as in Table II and the COD information is assumed to be unavailable.

Note that the  $g$  groups are in order of increasing dose levels  $0 = d_1 < d_2 < \dots < d_g$ , and denote  $\hat{\boldsymbol{p}} = (\hat{p}_1, \hat{p}_2, \dots, \hat{p}_g)$ . Suppose that the set of values  $\boldsymbol{p} = (p_1, p_2, \dots, p_g)$  consists of the binomial parameters indicating the actual probabilities of a positive response for each dose level. Under the monotonicity constraint of the dose-response relationship (i.e.,  $p_1 \leq p_2 \leq \dots \leq p_g$ ), our objective is to test the following hypotheses:

$$H_0: p_1 = p_2 = \dots = p_g$$

$$H_1: p_1 \leq p_2 \leq \dots \leq p_g \text{ with at least one strict inequality.}$$

## 3. ISOTONIC MODIFICATION OF THE COCHRAN-ARMITAGE TYPE TREND TESTS

### 3.1. The Cochran-Armitage (CA) Trend Test

Cochran<sup>1</sup> and Armitage<sup>2</sup> introduced a test for detecting a linear trend across dose groups in lifetime tumor incidence rates. This test needs an assumption that all animals are at equal risk of developing a tumor over the duration of study. Using the notation in Table II, the

expected number of animals with tumors in the  $i^{\text{th}}$  group is  $E_i = xK_i$ , where  $K_i = n_i/n$ . Defining  $D_i$  as  $x_i - E_i$ , the test statistic for possible monotonic trend with dose is based on

$$X = \sum_{i=1}^g d_i D_i,$$

and the variance is estimated by

$$V = \frac{x(n-x)}{n(n-1)} \sum_{i=1}^g n_i (d_i - \bar{d})^2, \quad (1)$$

where  $\bar{d} = \sum_{i=1}^g n_i d_i / n$ . The CA test statistic is  $Z_{\text{CA}} = X/\sqrt{V}$ , where  $Z_{\text{CA}}$  is asymptotically distributed as a standard normal variate under the null hypothesis of equal tumor incidence rates among the groups.

### 3.2. The Poly-3 Test (Bieler and Williams)

Bailer and Portier<sup>3</sup> 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. The weighting scheme is described below.

Bailer and Portier<sup>3</sup> define the number at risk for the  $i^{\text{th}}$  group as the sum of  $n_i$  weights:

$$r_i = \sum_{j=1}^{n_i} w_{ij}, \quad (2)$$

where  $w_{ij}$  is the time-at-risk weight for the  $j^{\text{th}}$  group. That is, the number of animals  $n_i$  in (1) is replaced with  $r_i$  in the CA test. The risk weight  $w_{ij}$  is defined as

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

where  $t_{ij} (\leq t_{\max})$  is the actual death time of the animal and  $t_{\max}$  is the time to termination of the experiment (e.g., terminal sacrifice time). Bailer and Portier recommend that  $k = 3$  be used since previous work conducted by Portier, Hedges, and Hoel<sup>6</sup> in fitting Weibull hazards

to tumor onset data showed that most tumors occur at the estimated rate of a third- to fifth-order polynomial in time. Note that this weighting gives less weight to a tumor-free animal that dies at time  $t_{ij}$ .

Bieler and Williams<sup>4</sup> suggested a further modification of the CA test using the delta method and weighted least squares techniques to adjust the variance estimation of ratio statistics. With the notation introduced earlier, define:

$$\hat{p}'_i = x_i/r_i \text{ and } \hat{p}' = x/\sum_i r_i ,$$

$$\hat{q}'_i = 1 - \hat{p}'_i \text{ and } \hat{q}' = 1 - \hat{p}' ,$$

$$a_i = r_i^2/n_i , \quad v_{ij} = x_{ij} - \hat{p}'_i w_{ij} , \text{ and } \bar{v}_i = \sum_j v_{ij}/n_i ,$$

where  $x_{ij}$  is the tumor status (yes/no) for the  $j^{\text{th}}$  animal in the  $i^{\text{th}}$  group ( $\sum_j x_{ij} = x_i$ ).

The test statistic of Bieler and Williams is

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

where  $C = \left[ \sum_i \sum_j (v_{ij} - \bar{v}_i)^2 \right] / (n - g)$ , and  $g$  is the number of groups.

### 3.3. Isotonic Modification and Bootstrap Method

The term “isotonic” is synonymous with “non-decreasing.” Refer to the data summary in Table II. Suppose that the  $p_i$ 's are isotonic, then the  $\hat{p}_i$ 's can be expected to be isotonic if the  $n_i$ 's are large enough. However, this is not always the case in practice. Even when the true dose-response curve is monotone, observing non-monotonic  $\hat{p}_i$ 's may occur. In this case, a conventional test procedure may not be able to detect a linear trend across dose groups due to the non-monotonicity in the observed proportions.

We propose an isotonic modification of the Cochran-Armitage type (CA and Poly-3) trend tests. The pool-adjacent-violators algorithm (PAVA), first published by Ayer et al.<sup>9</sup>,

is adopted here. First we summarize the data as in Table II. If the set of values  $\hat{p} = (\hat{p}_1, \hat{p}_2, \dots, \hat{p}_g)$  is isotonic, then the usual CA or Poly-3 test statistic is computed; otherwise, there must exist a neighboring pair  $(i, i + 1)$  such that  $\hat{p}_i > \hat{p}_{i+1}$ . We find the smallest such index, say  $i$ , and combine the two groups  $i$  and  $i + 1$  into a single group whose dose is the smaller one of the two dose levels and group sample size is the sum of the two. The observed proportion of this new group is then

$$\frac{x_i + x_{i+1}}{n_i + n_{i+1}},$$

the mean of the combined data. If this new set of the  $g - 1$  values is not isotonic, then we repeat this process until obtaining a data set with isotonic proportions. We refer to this iterative pairwise averaging process as “amalgamation.” The test statistic for the CA test or Poly-3 test is then to be calculated based on the resulting (combined) data.

The bootstrap method is applied in finding the empirical distributions of the test statistics and the corresponding critical values in the isotonic versions of the CA-type tests. For a data set,  $B$  bootstrap samples are generated. With the notation  $X$  and  $X^*$  denoting the original sample and the bootstrap sample respectively, the bootstrap process is described below:

1. A data set  $X$ , consisting of information on  $n$  animals across the  $g$  dose groups, is used to calculate the test statistic  $S(X)$ .
2.  $B$  bootstrap samples  $X^{*1}, X^{*2}, \dots, X^{*B}$  are generated from the original sample  $X$ . Each bootstrap sample has  $n$  elements, generated by sampling with replacement from the original data set  $X$ . It should be noted that each of the  $n$  animals in  $X$  is equally likely to be chosen (and may be chosen more than once) in spite of which dose group it is from, and the chosen animals are randomly assigned to each of the  $g$  groups (with  $n_i$  animals in group  $i$ ) for conducting a trend test. Bootstrap replicates  $S(X^{*1}), S(X^{*2}), \dots, S(X^{*B})$  are then obtained by calculating the value of the test statistic on each bootstrap sample.
3. The critical value  $CR(X)$  at the significance level  $\alpha$  is estimated by the  $100(1 - \alpha)^{\text{th}}$  percentile of the values  $S(X^{*1}), S(X^{*2}), \dots, S(X^{*B})$ .

4. Compare the test statistic  $S(X)$  obtained from the original data set with the critical value  $CR(X)$ . If  $S(X) \geq CR(X)$ , then  $H_0$  is rejected.

Alternatively, the empirical  $p$ -value can be obtained from the distribution of the bootstrap replicates.

## 4. ORDER-RESTRICTED LIKELIHOOD RATIO TEST WITH A SURVIVAL ADJUSTMENT

### 4.1. The Likelihood-Ratio Chi-Squared Test for Homogeneity

The hypotheses in the likelihood-ratio chi-squared test for homogeneity are:

$$H_0: p_1 = p_2 = \cdots = p_g$$

$H_1$ : otherwise.

The test is based on the ratio of the two probabilities: the maximized likelihood under  $H_0$  and the maximized likelihood under  $H_0 \cup H_1$ . Without constraint, the maximized likelihood is obtained at  $\hat{\mathbf{p}} = (\hat{p}_1, \hat{p}_2, \cdots, \hat{p}_g)$ , the set of observed proportions of positive responses. The  $\hat{p}_i$ 's are known as the maximum likelihood estimates (MLEs) of  $p_i$ 's. The maximized likelihood restricted to  $H_0$  is obtained at  $\hat{p}_0 = \sum_i x_i/n_i = x/n$ , the overall proportion. Let  $\Lambda$  denote the ratio of the maximized likelihoods, which cannot exceed 1. The test statistic is

$$-2 \log \Lambda = 2 \sum_{i=1}^g \left[ x_i \log \frac{\hat{p}_i}{\hat{p}_0} + (n_i - x_i) \log \frac{1 - \hat{p}_i}{1 - \hat{p}_0} \right],$$

which has an asymptotic chi-squared distribution with  $g - 1$  degrees of freedom under  $H_0$ .<sup>7,8</sup>

### 4.2. Order-Restricted Likelihood Ratio Test (LRT)

The standard MLE of  $\mathbf{p}$  is  $\hat{\mathbf{p}}$ . A monotonic MLE of  $\mathbf{p}$  is needed for conducting the order-restricted likelihood ratio test of Robertson, Wright and Dykstra.<sup>5</sup> Under the constraint that the event incidence rates are isotonic, the PAVA is used for computing  $\hat{\mathbf{p}}^* = \{\hat{p}_i^* : i = 1, 2, \cdots, g\}$ , the set of monotonic MLEs of  $p_i$ 's. The algorithm proceeds as follows.

PAVA starts with  $\hat{\boldsymbol{p}}$ . If  $\hat{\boldsymbol{p}}$  is isotonic, then  $\hat{\boldsymbol{p}}^* = \hat{\boldsymbol{p}}$ ; otherwise, there must be an index  $k$  such that  $\hat{p}_k > \hat{p}_{k+1}$ . Such elements  $\hat{p}_k$  and  $\hat{p}_{k+1}$  are called *violators*. The two adjacent violators  $\hat{p}_k$  and  $\hat{p}_{k+1}$  are then replaced by their weighted average  $AV_{k,k+1} = (x_k + x_{k+1}) / (n_k + n_{k+1})$ . The two weights  $n_k$  and  $n_{k+1}$  are replaced by  $n_k + n_{k+1}$ . If this new set of  $g - 1$  values is isotonic, then  $\hat{p}_k^* = \hat{p}_{k+1}^* = AV_{k,k+1}$  and  $\hat{p}_i^* = \hat{p}_i$  for  $i$  other than  $k$  or  $k + 1$ . If this new set of  $g - 1$  values is not isotonic, then this process is repeated using the new values and weights until an isotonic set of values is obtained. The resulting set  $\hat{\boldsymbol{p}}^*$  provides the MLE for  $\boldsymbol{p}$  under the order restriction.<sup>5</sup>

The algorithm determines a partition of the dose set  $\boldsymbol{d} = (d_1, d_2, \dots, d_g)$  into sets of consecutive elements on which the value of the monotonic MLE is constant. These sets are called the *level sets*. In testing the hypotheses  $H_0 : p_1 = p_2 = \dots = p_g$  against  $H_1 : p_1 \leq p_2 \leq \dots \leq p_g$  with at least one strict inequality, the test statistic based on the LRT is

$$T = 2 \sum_{i=1}^g \left[ x_i \log \frac{\hat{p}_i^*}{\hat{p}_0} + (n_i - x_i) \log \frac{1 - \hat{p}_i^*}{1 - \hat{p}_0} \right].$$

The asymptotic distribution of  $T$  under  $H_0$  is

$$\Pr(T > c) = \sum_{j=2}^g \left[ l(j, g) \Pr(\chi_{j-1}^2 > c) \right],$$

where the level probability  $l(j, g)$  denotes the probability that given  $g$  groups under  $H_0$  the isotonic regression will result in  $j$  level sets.<sup>5</sup> Note that  $\sum_{j=1}^g l(j, g) = 1$ .

### 4.3. Survival Adjustment

The presence of treatment-induced mortality unrelated to the tumor of interest may mislead tests that only focus on the crude lifetime tumor incidence rates. To address this issue in the LRT, we adopt the weighting scheme given in Equation (3) and propose a survival-adjusted LRT. This adjustment can be applied to the LRT for either the order-restricted or unrestricted parameter cases.

The number at risk for the  $i^{\text{th}}$  group as defined in Equation (2),  $r_i$ , is effectively used

in replacing  $n_i$  in the standard MLEs of  $p_i$ 's. That is, the standard survival-adjusted MLE of  $p_i$  becomes  $\hat{p}_i = x_i/r_i$  ( $i = 1, 2, \dots, g$ ). These survival-adjusted MLEs are then used in conducting the LRT.

On the other hand, in the case of the isotonic LRT, the monotonic MLEs of the  $p_i$ 's are obtained by applying PAVA in the following fashion while pooling two adjacent violators:

$$\hat{p}_i^* = \hat{p}_{i+1}^* = \frac{x_i + x_{i+1}}{r_i + r_{i+1}}.$$

Then the isotonic LRT is conducted using these monotonic survival-adjusted MLEs.

## 5. SIMULATION STUDY

### 5.1. Design of Monte Carlo Simulation

A Monte Carlo simulation study was conducted to make comparisons among the CA-type and LRT tests and to determine the benefits of survival adjustments and isotonic modifications. A typical bioassay design with four dose groups of 50 animals each, and an experimental duration of 104 weeks, which is a normal term for a chronic study in rodents, is used in the study. The design was simulated to have a single terminal sacrifice at the end of the experiment as in the customary lifetime rodent bioassay. The four dose levels used are 0, 1, 2, and 4 for the control, low, intermediate and high doses, respectively.

The three independent variables  $T_1$  (the time to tumor onset),  $T_2$  (the potential time from tumor onset until death from the tumor) and  $T_3$  (the potential time until death from a competing risk) are generally used to model animal tumorigenicity data. Note that the results of the trend tests considered in this paper are independent of  $T_2$ , unlike the tests for which the cause of death information is utilized.<sup>10,11</sup>

The survival function for  $T_1$  was modeled as  $S(t) = \exp[-\delta_1(t/104)^{\delta_2}]$ , a two-parameter Weibull distribution. The value of  $\delta_2$  is set to 3, and the value of  $\delta_1$  is chosen such that the probability of tumor onset by the end of the experiment is either 5%, 15%, 25% or 35%.

The survival function for  $T_3$  is taken to be of the form given by Portier et al.<sup>6</sup>:  $Q(t) =$

$\exp[-(\gamma_1 t + \gamma_2 t^{\gamma_3})]$ , where  $\gamma_1 \geq 0, \gamma_2 \geq 0$ , and  $\gamma_3 \geq 0$ . With  $\gamma_1 = 10^{-4}$  and  $\gamma_2 = 10^{-16}$ , the value of  $\gamma_3$  is chosen such that the competing risks survival rate (CRSR) at the end of the experiment is either 0.5 for all the dose groups or 0.5, 0.4, 0.3 and 0.2 for the control, low, intermediate, and high dose groups, respectively.

For each configuration, 10,000 simulated data sets are generated and tested by various tests. In conducting the isotonic versions of the CA and Poly-3 tests, 5000 bootstrap samples are generated from each simulated data set in estimating the critical values. After the outer loop of 10,000 data sets  $X_1, X_2, \dots, X_{10,000}$  is finished, 10,000 critical values  $CR(X_1), CR(X_2), \dots, CR(X_{10,000})$  can be obtained as well as the frequency of the rejections. The descriptive statistics including the mean, standard deviation, and quartiles for the values of the  $CR$ 's can be computed along with the empirical 95% confidence interval (i.e., the interval containing the middle 95% of the values of  $CR(X_1), CR(X_2), \dots, CR(X_{10,000})$ ).

## 5.2. Simulation Results

It is well-known that the age-adjusted tests are preferable under many circumstances. Our simulation study shows that the Poly-3 test performs better than the CA test, as expected. Similarly, the proposed survival-adjusted LRT shows great advantage over the unadjusted LRT. Therefore, only the results for the Poly-3 test (P3) and survival-adjusted LRT (S-LRT) are reported in Tables III and IV along with each of their isotonic versions. The size evaluation is displayed in Table III and the power evaluation is reported in Table IV.

For both standard and isotonic approaches, the P3 test controls size better than the S-LRT when the background tumor rate is as low as 0.05. The isotonic approach for P3 shows reduction of the probability of a Type I error, except for the low tumor rate 0.05 with different CRSR. The isotonic approach for S-LRT appears to control size for same CRSR and reduce size for different CRSR for high tumor rates 0.15, 0.25 and 0.35.

Although not reported, the CA test becomes too conservative for models with different CRSR as expected, and this also occurs in the unadjusted LRT due to the lack of survival adjustment. With the survival adjustment, S-LRT shows a significant improvement in controlling the size compared with LRT. Although the size is inflated for low tumor rate (.05),

the S-LRT maintains size quite well for higher tumor rates.

Simulating over a broad range of dose-response situations, the isotonic regression approach substantially increases the statistical power of the tests. In particular, the isotonic modification for the P3 test increases the power from .60 to .81 for the model with tumor rates (.05, .25, .25, .25) and same CRSR. Another example is in the case where the tumor rates are (.01, .05, .10, .15) with different CRSR. The power is improved by the isotonic approach for S-LRT from .62 to .84. The survival adjustment also results in higher power for the tests in situations where CRSR is different across dose groups. Although not reported, the Poly-3 test obtains greater power than the CA test in general. Similarly, the S-LRT appears to be more powerful than the unadjusted LRT.

The test statistics for the standard CA and Poly-3 tests are asymptotically distributed as standard normal under  $H_0$ . After the amalgamation process, the values of the test statistics for the isotonic CA and Poly-3 tests are non-negative. The bootstrap method is used in estimating the critical values of these tests. The distributions of these test statistics appear to be similar to  $|Z|$ , where  $Z \sim N(0, 1)$ .

## 6. EXAMPLE

The proposed methods are applied to analyze the C.I. Pigment Red 23 data discussed in the Introduction. In this example,  $t_{\max} = 729$  in Equation (3) was used since the terminal sacrifice started at the 729<sup>th</sup> day and lasted for a week. As shown in Figure 1, the observed dose-response curves for both TGC and TGF are non-monotonic. The count data before and after amalgamation are summarized in Table V. Table VI contains the results for all trend tests presented in this paper.

For TGC, the combined data resulting from amalgamation contain two groups of 100 animals each with dose levels 0 and 25,000 ppm. All of the CA, P3, LRT, and S-LRT show no statistical significance for trend at  $\alpha = .05$  while their isotonic counterparts attain an agreement in rejecting the null hypothesis of equal tumor incidence rates among the dose groups and show evidence of a toxic effect of C.I. Pigment Red 23 on development of C-cell

carcinoma of the thyroid gland.

The isotonic approach to the TGF data combines the middle two dose groups, and results in three groups of sample sizes (50, 100, 50) with dose levels (0, 10000, 50000). The isotonic CA/P3 reject the null hypothesis at  $\alpha = .05$ , consistent with the standard CA/P3. The LRT and S-LRT fail to reject the null hypothesis while their isotonic versions detect a significant linear trend in development of follicular-cell adenoma of the thyroid gland related to C.I. Pigment Red 23.

## 7. CONCLUDING REMARKS

In many animal bioassays for carcinogenicity, the prior monotonicity of dose-response relationship is assumed since higher doses are expected to produce stronger treatment effects within a reasonable dose range. The present paper develops an isotonic approach to obtain order-restricted Cochran-Armitage type trend tests. This approach can improve the efficiency of statistical analysis by increasing the power of the tests and facilitating the interpretation of the outcomes.

Another problem for the statistical analysis of tumor incidence rates arises from the presence of treatment-induced mortality. A standard LRT for trend could be biased since it focuses on crude lifetime tumor incidence rates and makes no adjustment for differences in survival experiences across dose groups. This survival adjustment can be adapted to the LRT for either the order-restricted or unrestricted parameter cases.

The Monte Carlo simulation study presented in this paper comparing these trend tests under various situations shows the advantages of the isotonic modifications and survival adjustment. Under many circumstances, our results indicate that the isotonic Poly-3 test is overall more favorable than others in terms of gaining statistical power while controlling the probability of a Type I error. In general, if an order-restricted trend test is required for analyzing animal tumorigenicity data in the absence of cause-of-death information, our recommendation is to apply the isotonic Poly-3 test. The isotonic version of the survival-adjusted LRT can serve as an alternative method.

## ACKNOWLEDGEMENTS

Jessica Chang and Hongshik Ahn's work was supported by NIH grant R29 CA77289-03. Hongshik Ahn's research was partially supported by the Faculty Research Participation Program at the National Center for Toxicological Research administered by the Oak Ridge Institute for Science and Education through an interagency agreement between USDOE and USFDA.

## REFERENCES

1. Cochran, W. G. 'Some methods for strengthening the common  $\chi^2$  tests', *Biometrics*, **10**, 417-451 (1954).
2. Armitage, P. 'Tests for linear trends in proportions and frequencies', *Biometrics*, **11**, 375-386 (1955).
3. Bailer, A. J. and Portier, C. J. 'Effects of treatment-induced mortality and tumor-induced mortality on tests for carcinogenicity in small samples', *Biometrics*, **44**, 417-431 (1988).
4. Bieler G. S. and Williams R. L. 'Ratio estimates, the delta method, and quantal response tests for increased carcinogenicity', *Biometrics*, **49**, 793-801 (1993).
5. Robertson, T., Wright, F. T., and Dykstra, R. L. *Order-Restricted Statistical Inference*, New York: John Wiley, 1988.
6. Portier, C., Hedges, J. and Hoel, D. G. 'Age-specific models of mortality and tumor onset for historical control animals in the National Toxicology Program's carcinogenicity experiments', *Cancer Research*, **46**, 4372-4378 (1986).
7. Wilks, S. S. 'The likelihood test of independence in contingency tables', *Annals of Mathematical Statistics*, **6**, 190-196 (1935).
8. Wilks, S. S. 'The large-sample distribution of the likelihood ratio for testing composite hypotheses', *Annals of Mathematical Statistics*, **9**, 60-62 (1938).

9. Ayer, M., Brunk, H. D., Ewing G. M., Reid, W. T., and Silverman, E. 'An empirical distribution function for sampling with incomplete information', *Annals of Mathematical Statistics*, **26**, 641-647 (1955).
10. Peto, R. 'Guidelines on the the analysis of tumor rates and death rates in experimental animals', *British Journal of Cancer*, **29**, 101-105 (1974).
11. Peto, R., Pike, M. C., Day, N. E., Gray, R. G., Lee, P. N., Parish, S., Peto, J., Richards, S. and Wahrendorf, J. '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 (1980). International Agency for Research on Cancer: Lyon, France.

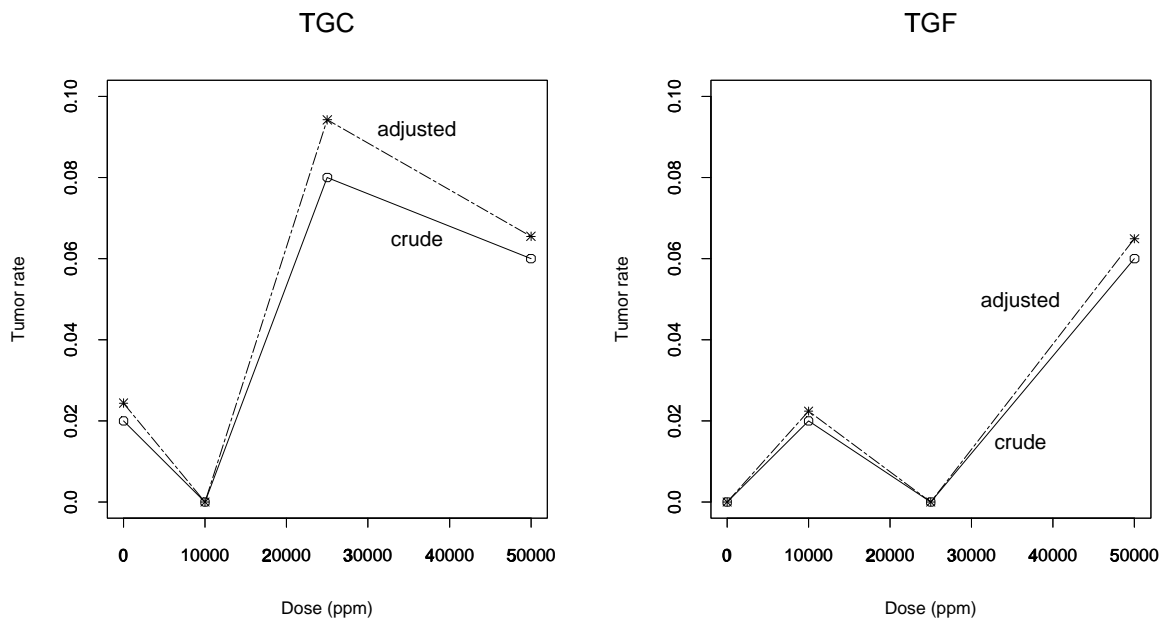


Figure 1: Crude and survival-adjusted tumor proportions in the C.I. Pigment Red 23 data for each of the C-cell carcinoma (TGC) and follicular cell adenoma (TGF) of the thyroid gland



Table II: Tumor data summary

	Dose Group				Total
	1	2	$\cdots$	$g$	
# with tumors	$x_1$	$x_2$	$\cdots$	$x_g$	$x$
# of objects	$n_1$	$n_2$	$\cdots$	$n_g$	$n$
Proportions	$\hat{p}_1 = \frac{x_1}{n_1}$	$\hat{p}_2 = \frac{x_2}{n_2}$	$\cdots$	$\hat{p}_g = \frac{x_g}{n_g}$	$\hat{p}_0 = \frac{x}{n}$

Table III: Simulated size corresponding to nominal 5% significance level for Poly-3, survival-adjusted LRT, and their isotonic versions. All tests are based on a dose scaling of 0, 1, 2, 4 with 10,000 trials for each configuration. The critical values for the isotonic Poly-3 test are based on 5,000 bootstrap samples

TR <sup>a</sup>	CRSR <sup>b</sup>	Approach	P3	S-LRT
0.05	same <sup>c</sup>	standard	0.0601	0.0715
		isotonic	0.0553	0.0738
	diff <sup>d</sup>	standard	0.0526	0.0691
		isotonic	0.0564	0.0729
0.15	same	standard	0.0562	0.0565
		isotonic	0.0499	0.0517
	diff	standard	0.0484	0.0548
		isotonic	0.0472	0.0487
0.25	same	standard	0.0556	0.0454
		isotonic	0.0520	0.0495
	diff	standard	0.0459	0.0508
		isotonic	0.0433	0.0427
0.35	same	standard	0.0486	0.0440
		isotonic	0.0461	0.0447
	diff	standard	0.0438	0.0433
		isotonic	0.0430	0.0381

<sup>a</sup>Cumulative tumor onset probability at 104 weeks in absence of competing risks

<sup>b</sup>Competing Risks Survival Rate with respect to all causes except for the tumor of interest

<sup>c</sup>0.5 for all dose groups

<sup>d</sup>0.5, 0.4, 0.3, 0.2 for control, low, intermediate and high dose groups, respectively

Table IV: Simulated power corresponding to nominal 5% significance level for Poly-3, survival-adjusted LRT, and their isotonic versions. All tests are based on a dose scaling of 0, 1, 2, 4 with 10,000 trials for each configuration. The critical values for isotonic Poly-3 test are based on 5,000 bootstrap samples

Model <sup>a</sup>	CRSR <sup>b</sup>	Approach	P3	S-LRT
.05 .05 .05 .25	same <sup>c</sup>	standard	0.9289	0.8253
		isotonic	0.9308	0.9150
	diff <sup>d</sup>	standard	0.8590	0.7478
		isotonic	0.8691	0.8510
.05 .05 .25 .25	same	standard	0.9283	0.9037
		isotonic	0.9646	0.9673
	diff	standard	0.9142	0.8653
		isotonic	0.9484	0.9498
.05 .15 .25 .25	same	standard	0.7983	0.7321
		isotonic	0.8521	0.8937
	diff	standard	0.7977	0.7002
		isotonic	0.8389	0.8724
.05 .25 .25 .25	same	standard	0.5988	0.7769
		isotonic	0.8093	0.9065
	diff	standard	0.6288	0.7498
		isotonic	0.8170	0.8902
.05 .15 .25 .35	same	standard	0.9779	0.9124
		isotonic	0.9715	0.9839
	diff	standard	0.9668	0.8764
		isotonic	0.9531	0.9649
.01 .05 .10 .15	same	standard	0.8107	0.6733
		isotonic	0.7890	0.8714
	diff	standard	0.7882	0.6229
		isotonic	0.7632	0.8381

<sup>a</sup>Cumulative tumor onset probability at 104 weeks in absence of competing risks

<sup>b</sup>Competing Risks Survival Rate with respect to all causes except for the tumor of interest

<sup>c</sup>0.5 for all dose groups

<sup>d</sup>0.5, 0.4, 0.3, 0.2 for control, low, intermediate and high dose groups, respectively

Table V: Data summary for the C.I. Pigment Red 23 experiment

Tumor Test		Amalgamation Process						
		Before				After		
TGC <sup>a</sup>	Dose	0	10,000	25,000	50,000	0	25,000	
	CA # with tumor	1	0	4	3	1	7	
	$n_i$ <sup>c</sup>	50	50	50	50	100	100	
	P3 # with tumor	1	0	4	3	1	7	
	$r_i$ <sup>d</sup>	41.01	44.70	42.45	45.81	85.71	88.26	
TGF <sup>b</sup>	Dose	0	10,000	25,000	50,000	0	10,000	50,000
	CA # with tumor	0	1	0	3	0	1	3
	$n_i$	50	50	50	50	50	100	50
	P3 # with tumor	0	1	0	3	0	1	3
	$r_i$	41.01	44.70	42.45	46.20	41.01	87.15	46.20

<sup>a</sup>C-cell carcinoma of the thyroid gland

<sup>b</sup>Follicular cell adenoma of the thyroid gland

<sup>c</sup>  $n_i$ : sample size for group  $i$

<sup>d</sup>  $r_i$ : number at risk for group  $i$  as defined in Equation (2)

Table VI: Values of test statistics with  $p$ -values in parentheses for the C.I. Pigment Red 23 data

Tumor	Approach	CA	P3	LRT	S-LRT
TGC <sup>a</sup>	standard	1.529 (0.063)	1.448 (0.074)	6.800 (0.079)	6.833 (0.077)
	isotonic	2.160 (0.039)	2.147 (0.042)	5.250 (0.035)	5.096 (0.037)
TGF <sup>b</sup>	standard	2.006 (0.022)	1.975 (0.024)	6.715 (0.082)	6.322 (0.097)
	isotonic	2.361 (0.015)	2.308 (0.016)	5.319 (0.033)	4.976 (0.040)

<sup>a</sup>C-cell carcinoma of the thyroid gland

<sup>b</sup>Follicular cell adenoma of the thyroid gland