A Computational Tool for Testing Dose-related Trend Using an Age-adjusted Bootstrap-based Poly-k Test

Hojin Moon
NCTR, US FDA

Hongshik Ahn
The Stony Brook University, NY

Ralph L. Kodell
NCTR, US FDA

Abstract

A computational tool for testing for a dose-related trend and/or a pairwise difference in the incidence of an occult tumor via an age-adjusted bootstrap-based poly-k test and the original poly-k test is presented in this paper. The poly-k test (Bailer and Portier 1988) is a survival-adjusted Cochran-Armitage test, which achieves robustness to effects of differential mortality across dose groups. The original poly-k test is asymptotically standard normal under the null hypothesis. However, the asymptotic normality is not valid if there is a deviation from the tumor onset distribution that is assumed in this test. Our age-adjusted bootstrap-based poly-k test assesses the significance of assumed asymptotic normal tests and investigates an empirical distribution of the original poly-k test statistic using an age-adjusted bootstrap method. A tumor of interest is an occult tumor for which the time to onset is not directly observable. Since most of the animal carcinogenicity studies are designed with a single terminal sacrifice, the present tool is applicable to rodent tumorigenicity assays that have a single terminal sacrifice. The present tool takes input information simply from a user screen and reports testing results back to the screen through a user-interface. The computational tool is implemented in C/C++ and is applied to analyze a real data set as an example. Our tool enables the FDA and the pharmaceutical industry to implement a statistical analysis of tumorigenicity data from animal bioassays via our age-adjusted bootstrap-based poly-k test and the original poly-k test which has been adopted by the National Toxicology Program as its standard statistical test.

Keywords: bioassay, competing risks, likelihood, sacrifice, survival-adjusted test.

1. Introduction

Animal carcinogenicity bioassays are routinely used to evaluate the carcinogenic potential of chemical substances to which humans are exposed. Standard long-term animal carcinogenicity

studies of pharmaceuticals and food additives are usually conducted in both sexes of mice and rats for the majority of those animals' typical life spans. The Center for Drug Evaluation and Research (CDER) in the US Food and Drug Administration (FDA) recommends that drug sponsors conduct carcinogenicity studies at least 18 months in mice and 24 months in rats (Office of the Federal Register 1985). Kodell, Lin, Thorn, and Chen (2000) studied the effect of shortened duration on the statistical power of carcinogenicity studies, and the results support the CDER recommendation. Each animal is assumed to begin with a tumor free state. Animals are randomized into a control group (typically, animals that are exposed to a control agent or observed without any exposure) or into 2 to 3 treatment groups that receive specified levels of exposure until they either die or are sacrificed. Most of the animal carcinogenicity studies are designed with a single terminal sacrifice. In a single terminal sacrifice, all surviving animals are sacrificed and subjected to necropsy at the end of the experiment. During the study, age at death and the information on the presence or absence of the tumor of interest are collected for each animal. The primary goal of this type of experiment is to assess a dose-related trend of test agent exposures on the incidence of an occult tumor for which the time to tumor onset is not directly observable. Ahn and Kodell (1998) review various statistical testing schemes for a dose-response relationship.

The logrank test of Mantel and Haenszel (1959) may be used for comparing hazards of death from rapidly lethal tumors. To compare the prevalence of nonlethal tumors, the prevalence test proposed by Hoel and Walburg (1972) may be used for incidental tumors. However, the data obtained from a carcinogenicity experiment generally contain a combination of fatal and incidental tumors. Peto, Pike, Day, Gray, Lee, Parish, Peto, Richards, and Wahrendorf (1980) suggested combining the fatal and incidental tests, known as the cause-of-death (COD) test or the Peto test, for comparing tumor onset distributions.

The Cochran-Armitage (CA) test (Cochran 1954; Armitage 1955, 1971) was introduced for detecting a linear trend across dose groups in tumor incidence rates. This test does not require COD information. It assumes under the null hypothesis that all animals are at equal risk of developing the tumor among the dose groups over the duration of the study. A problem for this test arises from the presence of treatment-induced mortality unrelated to the tumor of interest. The CA test is sensitive to these changes in treatment lethality (Bailer and Portier 1988; Kodell, Chen, and Moore 1994; Moon, Ahn, and Kodell 2002, 2005; Moon, Ahn, Kodell, and Lee 2003).

The poly-k test (Bailer and Portier 1988) is a survival-adjusted quantal-response procedure that modifies the CA test to take dose-group differences in intercurrent mortality into account. Bieler and Williams (1993) further modified the poly-3 test by an adjustment of the variance estimation of the test statistic using the delta method (Woodruff 1971).

The statistical analysis of animal carcinogenicity data, including controversy surrounding the Peto COD test, has been addressed recently in the government-regulated pharmaceutical industry (Lee, Fry, Fairweather, Haseman, Kodell, Chen, Roth, Soper, and Morton 2002; STP Peto Analysis Working Group 2001; US Food and Drug Administration 2001). An appropriate alternative to the Peto-type test is the poly-3 test, which has been adopted by the National Toxicology Program (NTP) as its official test for carcinogenicity. The poly-3 procedure does not require COD data but does require an assumption about the shape of the Weibull-family tumor onset distribution.

The optimal value of k in the poly-k test relies on the shape of the tumor incidence function.

As Bailer and Portier mentioned, if the shape of the tumor incidence function is expected to follow time to some power k, which is different from 3, then the poly-k test with $k \neq 3$ should have superior operating characteristics to the poly-3 test. Recently, Moon *et al.* (2003) proposed a method for estimating k in the poly-k test. Later, Moon *et al.* (2005) developed an age-adjusted bootstrap-based poly-k test instead of directly estimating k in the poly-k test. The main idea of the age-adjusted bootstrap-based poly-k test developed by Moon *et al.* (2005) was to replace z_α from the normal 5 percent significance level with a bootstrap critical value for the poly-k test via the proposed age-adjusted bootstrap method instead of directly estimating k in the poly-k test. In this age-adjusted poly-k test, not only was the tumor incidence rate under H_0 preserved but also the competing risks survival rate was not altered.

The purpose of this paper is to present a user-friendly statistical tool for testing a dose-related trend using the age-adjusted bootstrap-based poly-k test developed by Moon *et al.* (2005) and the original poly-k test of Bailer and Portier (1988) with the modification developed by Bieler and Williams (1993). Our software can be used to statistically analyze a dose-related trend or to make a pairwise comparison of test agent versus control in rodent bioassays. Users can analyze animal tumorigenicity data efficiently using our simple and user-friendly statistical testing tool. A comprehensive list of inputs can be specified by the user through a simple input text file. The poly-k test is described in Section 2. Our age-adjusted bootstrap-based poly-k test is described in Section 3. A detailed description of the usage of the proposed tool is demonstrated in Section 4 with a real example from an animal tumorigenicity experiment. Concluding remarks are described in the last section.

2. The poly-k survival-adjusted trend test

The Cochran-Armitage (CA) test utilizes the tumor data pooled over the entire duration of a study for each group. The tumor data can be summarized as follows:

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

Under the null hypothesis, the expected number of animals with tumors in the i-th group is $E_i = yN_i/N$. The test statistic for a possible monotonic trend with dose is based on

$$X = \sum_{i=1}^{g} d_i(y_i - E_i),$$

and the variance is estimated by

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

where $\bar{d} = \sum_{i=1}^{g} N_i d_i/N$. The CA test statistic is

$$Z_{\mathrm{CA}} = X/\sqrt{V},$$

where Z_{CA} is asymptotically distributed as a standard normal variate under the null hypothesis of equal tumor incidence rates among the groups. Despite the assumption of equal risk of getting the tumor during the study, some treatments shorten overall survival so that decreased risks of tumor onset may occur. The CA test is known to be sensitive to increases in treatment lethality. In such a case, the CA test often fails to control the probability of a Type I error (Bailer and Portier 1988; Mancuso, Ahn, Chen, and Mancuso 2002; Moon *et al.* 2002, 2003, 2005).

Bailer and Portier (1988) proposed the poly-k test, which made an adjustment of the CA test by using a fractional weighting scheme for animals according to information on tumor presence/absence at the age of death. They define the number at risk for the i-th group as the sum of N_i weights

$$r_i = \sum_{h=1}^{N_i} w_{ih}, \tag{2}$$

where w_{ih} is the time-at-risk weight for the h-th animal in the i-th group. That is, the number of animals N_i in the CA test given in (1) is replaced with r_i, and $r = \sum_i r_i$ is used in place of N in (1). The risk weight w_{ih} is defined as

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

where t_{ih} ($\leq t_{\max}$) is the actual death time of the animal, and $t_{\max}$ is the time to termination of the experiment (e.g., terminal sacrifice time). This weighting scheme gives less weight to a tumor-free animal that dies at $t_{ih} < t_{\max}$.

Bieler and Williams (1993) 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

$$\hat{p}_i = \frac{y_i}{r_i} \text{ and } \hat{p} = \frac{y}{\sum_i r_i}; \quad a_i = r_i^2/N_i, \ v_{ih} = y_{ih} - \hat{p}w_{ih}, \text{ and } \bar{v}_i = \sum_h v_{ih}/N_i,$$

where y_{ih} is the tumor indicator (1 for presence and 0 for absence) for the h-th animal in the i-th group. The test statistic of Bieler and Williams is

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

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

3. Age-adjusted bootstrap-based method

The poly-k statistic is asymptotically standard normal under the null hypothesis of equal tumor incidence rates among the dose groups (Bieler and Williams 1993). This assumption is valid only if the correct value of k is used in the poly-k test. In order to find a suitable k for the poly-k test, Moon *et al.* (2003) proposed a method to estimate k for data with interval sacrifices.

Moon *et al.* (2005) recently developed the method of bootstrap resampling with an age-adjusted scheme as an alternative approach to estimate k for the poly-k test. We estimate the empirical distribution of the test statistic and the corresponding critical value of the poly-3 test while taking into account the presence of competing risks. It is accomplished via a modification of the permutation method of Farrar and Crump (1990) used in exact statistical tests. An approximately valid permutation procedure on the same general footing as the bootstrap method could also be developed. The goal of this study is to make the poly-3 test robust to the various Weibull-family tumor onset distributions and various competing risks survival rates in rodent bioassays with a single terminal sacrifice.

For a data set, B bootstrap samples will be generated as in Figure 1. With the notation X and X^* denoting the original sample (a data set) and the bootstrap sample respectively, the bootstrap resampling can be carried out as described in Algorithm 1. X and X^* are vectors for each animal containing death times of animals and tumor status across the G dose groups. A data set X is used to calculate the poly-3 statistic $T(X)$.

Algorithm 1 *Bootstrap Method*

1. *B bootstrap samples $X^{*1}, X^{*2}, \ldots, X^{*B}$ are generated from the pooled original sample X. Each bootstrap sample contains n elements, uniformly generated by sampling with replacement from the original data set X. The chosen animals are randomly assigned to each of the G groups for conducting the test. Then, $T(X^{*1}), T(X^{*2}), \ldots, T(X^{*B})$ are obtained by calculating the value of the test statistic on each bootstrap sample.*

2. *The critical value Z_α^B, a threshold for rejecting the null hypothesis of equal tumor incidence rate at the significance level α, is estimated by the $100(1-\alpha)$th percentile of the values $T(X^{*1}), T(X^{*2}), \ldots, T(X^{*B})$.*

3. *If $T(X) \geq Z_\alpha^B$, then the null hypothesis of equal tumor incidence rate across dose groups is rejected.*

The method described in Algorithm 1 is suitable for data with the same competing risks survival rate (CRSR). However, we need to note that if the CRSR is different across dose groups in the original data, the bootstrap samples from the pooled data may not reflect the CRSR of each group, while preserving the null distribution of equal tumor incidence rates across dose groups. In order to preserve the survival rates in each dose group, we need to modify the above bootstrap method and develop an age-adjusted bootstrap-based scheme (Figure 2). The age-adjusted bootstrap method developed by Moon *et al.* (2005) is illustrated in Algorithm 2.

Algorithm 2 *Age-adjusted Bootstrap Scheme*

1. *For the ith group, partition the total days of observation into $I(i, m), i = 1, \ldots, G; m = 1, \ldots, M_i$, consecutive intervals according to death times of animals in that group. These intervals need not correspond across groups, either in the number of intervals or in the number of days assigned to particular intervals. We denote by $A(I(i, m))$ the set of animals of the ith group whose death time is in the interval $I(i, m)$. We pool the animals that died or were sacrificed within the given interval, say $I(i, m)$, across dose groups. If this interval or interval $I(i, m + 1)$ does not contain any animal that died in other*
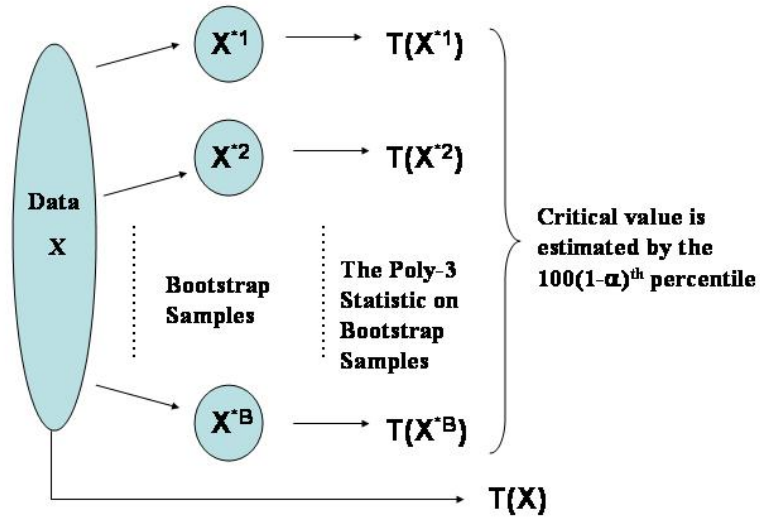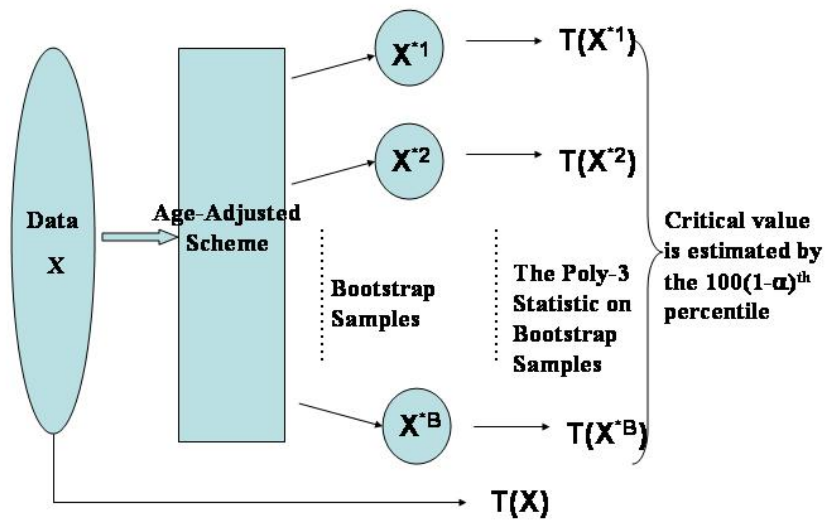
Figure 1: Bootstrap Method

Figure 2: Age-adjusted Bootstrap Scheme

groups, then these two intervals are merged. This merging process is continued until the interval contains at least one animal from another group.

2. *Intervalization for bootstrap is followed by pseudo-permutation of animals in each bootstrap-interval. Suppose the mth interval of the ith group contains r animals, and the mth interval contains R animals in total. For each bootstrap re-sample, we first pool the animals within this interval and shuffle them to randomly select r animals without replacement for the ith group. Suppose there are x tumor-death animals among those r animals. We then bootstrap r animals among the R animals in the mth interval. If the number of tumor deaths exceeds x in the bootstrap of r animals, we randomly select animals that died without tumor in that interval to replace the excess tumor-death animals. This step is required for bootstrap to preserve the mortality pattern in each group, while satisfying the null distribution of equal tumor incidence rate across dose groups. The B bootstrap samples, based on these strata, are generated as illustrated in Algorithm 1.*

Figure 3 illustrates the method for generating the first few animals in each group. In Group 1, the first two intervals in days are (0,185] and (185, 345]. In this group, animals (A–D) in the first interval are shuffled at each time of bootstrap to determine the maximum allowable number of tumor-death animals for the bth bootstrap sample, where $b = 1, \ldots, B$. Typically,
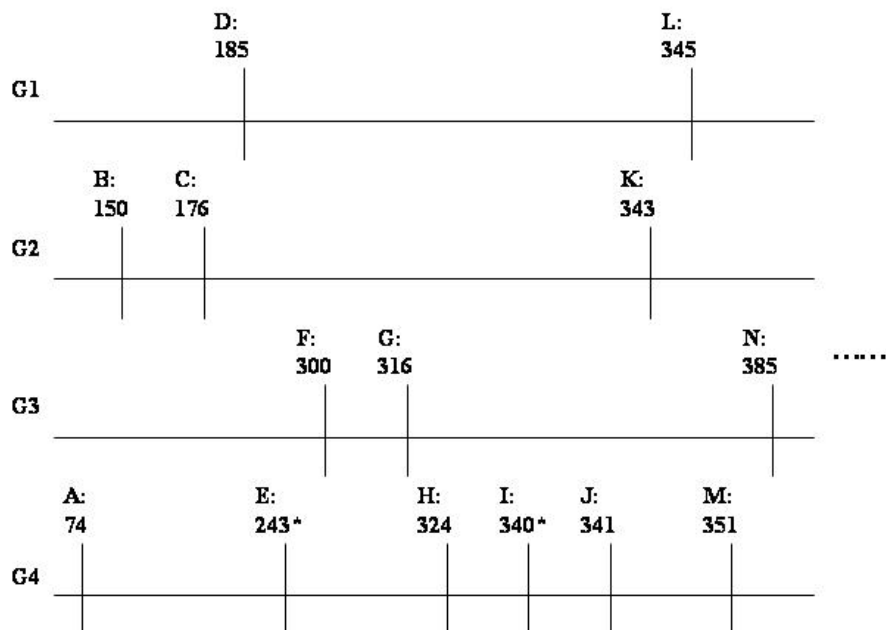

Figure 3: Death times (in days) in a hypothetical animal carcinogenicity data set with four groups (* indicates death with tumor; A – N indicates animal ID; Gx indicates the xth group).

the number of bootstrap samples should be at least 1000 for bootstrap confidence interval construction (Efron and Tibshirani 1993). We take $B = 5000$ bootstrap samples in this study. Then, the first bootstrap animal is generated randomly from the four animals (A–D) that died within interval (0,185]. If the maximum allowable tumor deaths for the bth bootstrap sample is zero, but the bootstrap animal died with tumor, then the animal is discarded, and an animal died without tumor among animals (A–D) is re-sampled. Otherwise, the bootstrap animal is kept. On the other hand, if the maximum allowable tumor death for the bth bootstrap sample is one, then the first bootstrap animal generated randomly from animals (A–D) is kept regardless of the tumor status of the first bootstrap animal. The second animal is generated randomly from the eight animals (E–L) in (185, 345] in the same fashion.

In Group 2, the first interval is chosen to be (0, 176] because there is no animal that died in other groups in (150, 176]. In this group, the first two animals are randomly sampled with replacement from the three animals (A–C) in (0, 176] up to but not more than the maximum allowable number of tumor animals determined in the shuffle, and the third animal is randomly generated from the eight animals (D–K) in (176, 343] in the same way. In Group 3, the first two animals are generated randomly from the seven animals (A–G) in (0, 316] with replacement, and the third animal is generated randomly from the seven animals (H–N) in (316, 385] in the same fashion. In Group 4, the first two animals are randomly generated from the five animals (A–E) in (0, 243] with replacement, and the next three animals are randomly generated from the five animals (F–J) in (243, 341] with replacement as explained above. The sixth animal is randomly generated from the four animals (K–M) in the same way, and so forth.

4. Usage of the age-adjusted poly-k testing software

A statistical tool for testing a dose-related trend and pairwise difference of exposures of a test agent on the incidence of an occult tumor via the age-adjusted bootstrap-based poly-k test of Moon *et al.* (2005) and the poly-k test of Bieler and Williams (1993) is developed. The computational tool is implemented in C/C++ and is freely available. Both, the source code and an executable Windows binary 'E7171.exe', are available with this paper. We will illustrate the usage of the age-adjusted poly-k testing tool on Windows PC using a data set from the 2-year gavage study of furan.

Toxicology and carcinogenesis studies were conducted by administering furan (C_4H_4O), a clear and colorless liquid, in corn oil by gavage to groups of F344/N rats and B6C3F$_1$ mice of each sex for two years (National Toxicology Program 1993). Furan was nominated by the National Cancer Institute for evaluation of carcinogenic potential due to its large production volume and use and because of the potential for widespread human exposure to a variety of furan-containing compounds. Corn oil was chosen as the vehicle for these studies. As an illustration, we focus on female F344/N rats for evaluation of carcinogenic potential on incidences of cholangiocarcinoma or hepatocellular neoplasms of the liver. The test results on the incidence of neoplasms from our proposed methods are compared to those from the conventional poly-3 test (Bieler and Williams 1993) which has been adopted by NTP. Groups of 50 rats of each sex were administered 2, 4 or 8 mg furan per kg body weight in corn oil by gavage 5 days per week for 2 years.

A simple format of an input file from the two-year gavage study of furan is illustrated as

follows:

```
50      0.0
470.0   2
528.0   2
537.0   2
...
728.0   4
728.0   4

50      2.0
364.0   2
436.0   2
478.0   2
...
728.0   4
728.0   4
```

It includes the total number of animals considered in the testing and the dose level for each group. The total number of groups should be greater than or equal to two for testing, including the control group. It is two for the pairwise test and greater than two for the trend test. In this example, there are two dose groups for testing (i.e., one for the control group and another for a dose group). A total of 100 animals are considered. Two dose levels are 0 mg furan per kg body weight for the control group, 2 mg/kg furan for the dose group. The corresponding data file '`twogpexdata.txt`' is available with this paper.

Two-column data immediately follow these input parameters. The first column indicates a death time (in days) for each animal. It should be between 0 and the terminal sacrifice time, inclusive. Typically, animal carcinogenicity studies on occult tumors are conducted for a duration of 104 weeks (728 days). The terminal sacrifice is scheduled at the end of the study. The last column is an indicator of mode of death (MOD) for tumor/nontumor and natural/sacrifice death (1 for natural death with tumor; 2 for natural death without tumor; 3 for sacrifice with tumor; 4 for sacrifice without tumor). For example, the 4th row of the data set in the control group (537.0, 2) indicates that an animal in the control group died naturally without tumor at day 537. It is easy to see that the scheduled time for terminal sacrifice is at 104 weeks (728 days).

To run the age-adjusted poly-k testing tool in Microsoft Windows OS (XP), a user can simply double-click the executable file '`E7171.exe`' in a directory where the execution file is stored. Although the package is written in Microsoft Visual C/C++ 6.0, a user is not required to have the Microsoft Visual C/C++ program to run the software. When a user runs the executable file '`E7171.exe`', the program starts and shows Figure 4.

Before running a program, a user needs to load an input data file, and needs to specify an output file by clicking '`Save Output`'. This saved output file may be printed if necessary. By clicking '`Load data`' button in Figure 5, a pop-up window will appear so that a user can go to a directory where an input file is located. Upon loading an input data file, the number of dose groups, total subjects, the number of subjects in each group and dose levels of each group will appear in the window. In Figure 5, two buttons are highlighted in '`Group`' buttons because the input data file includes two dose groups. If a user clicks one of the highlighted

Figure 4: The age-adjusted poly-*k* test: Before launching a data set for testing

Figure 5: The age-adjusted poly-*k* test: After launching a data set for testing

button '`1`' or '`2`' in '`Group`' button, death times and the mode of death for each subject will appear under '`Data`' section. To test the carcinogenicity activity of furan between two groups, a user may click '`Run`' button.

Figure 6 not only shows test results on the carcinogenicity activity of furan in female F344/N rats based on increased incidences of cholangiocarcinoma and hepatocellular neoplasms of the liver from the age-adjusted poly-k test, the poly-1 test, the poly-3 test and the poly-6 test, but also shows the observed proportions and the poly-3 adjusted proportions in each dose group. The two-group comparison of the vehicle control versus the 2 mg/kg dose group shows different conclusions between the age-adjusted poly-k test and the poly-3 test. In contrast with the result of the poly-3 test ($p = 0.0716$), our age-adjusted poly-k test shows significant carcinogenic activity of furan ($p = 0.0394$) at the level of 0.05. NTP concluded that under the conditions of these 2-year gavage studies, there was clear evidence of carcinogenic activity of furan in female F344/N rats based on increased incidences of cholangiocarcinoma and hepatocellular neoplasms of the liver. Results from our age-adjusted poly-k test applied to any combination of experimental groups agree with the conclusions of NTP. For example, Figure 7 shows the carcinogenicity activity of furan among three dose groups. It shows consistent results among our age-adjusted poly-k test, the poly-1 test, the poly-3 test and the poly-6 test. The corresponding data file '`threegpexdata.txt`' is also available with this paper.

5. Discussion

The importance of the poly-k test has been highlighted recently. In May 2001, FDA's Center for Drug Evaluation and Research (CDER) published in the Federal Register a guidance document (US Food and Drug Administration 2001). In that document, CDER endorsed the use of the Peto procedure for analyzing tumorigenicity data, although the poly-3 test was mentioned as a possible alternative when cause-of-death (COD) data (equivalently, context-of-observation data) are not available. The Society of Toxicologic Pathologists (STP) published in its journal a commentary on CDER's guidance that was critical of Peto's procedure and that largely endorsed the poly-3 procedure (STP Peto Analysis Working Group 2001). The poly-3 procedure does not require COD data but does require a critical assumption about the shape of the tumor onset distribution. Later, the Society withdrew its criticisms of the CDER guidance document on the Peto approach, while still recognizing the appropriateness of the poly-3 test in certain situations (STP Peto Analysis Working Group 2002).

It is clear that the statistical analysis of tumorigenicity data from animal bioassays remains an important regulatory issue to FDA and the pharmaceutical industry. We have developed a statistical tool using the age-adjusted poly-k test for animal carcinogenicity studies to detect a dose-related trend and/or a pairwise difference in tumor incidence following exposure to a putative carcinogen. It is applicable for studies on occult tumors for which the time to tumor onset is not directly observable.

As examples, application of this tool was illustrated with NTP data sets of the 2-year gavage study of furan. Use and comparison of our age-adjusted poly-k test and the poly-3 test of Bieler and Williams was made in the application to a data set of female F344/N rats on incidences of cholangiocarcinoma and hepatocellular neoplasms of the liver. Our software provides a benefit to FDA and pharmaceutical industry by furnishing a user-friendly tool for
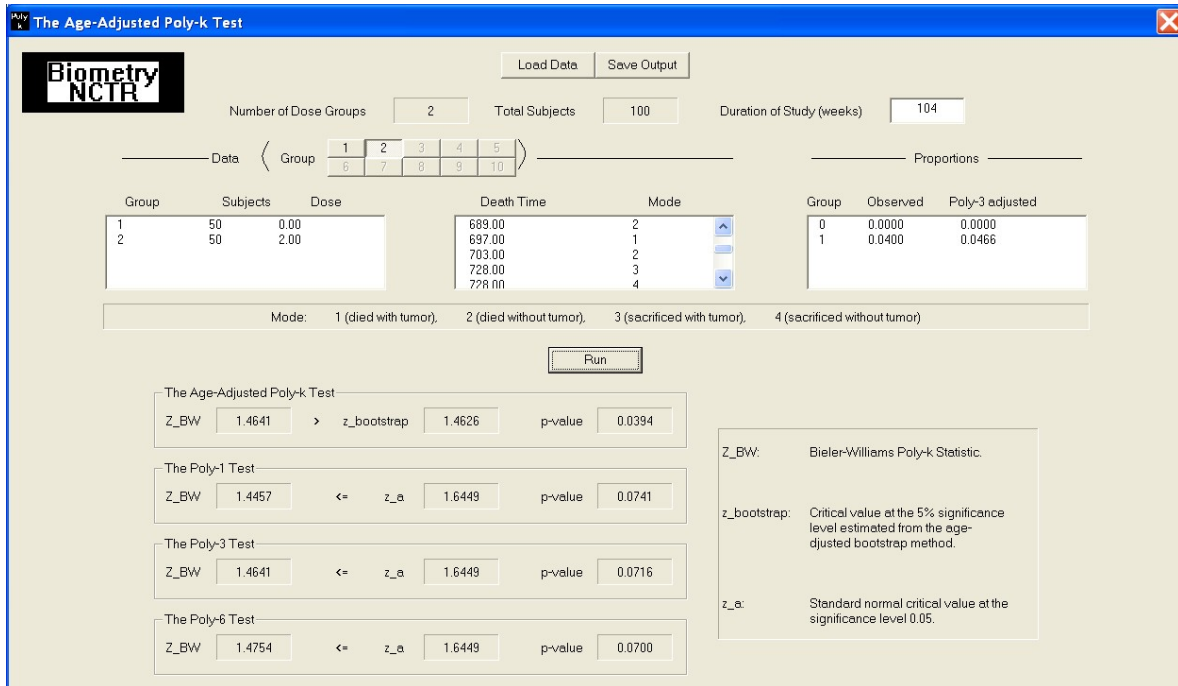
Figure 6: The age-adjusted poly-k test, the poly-1 test, the poly-3 test and the poly-6 test on the carcinogenicity activity of furan between two groups
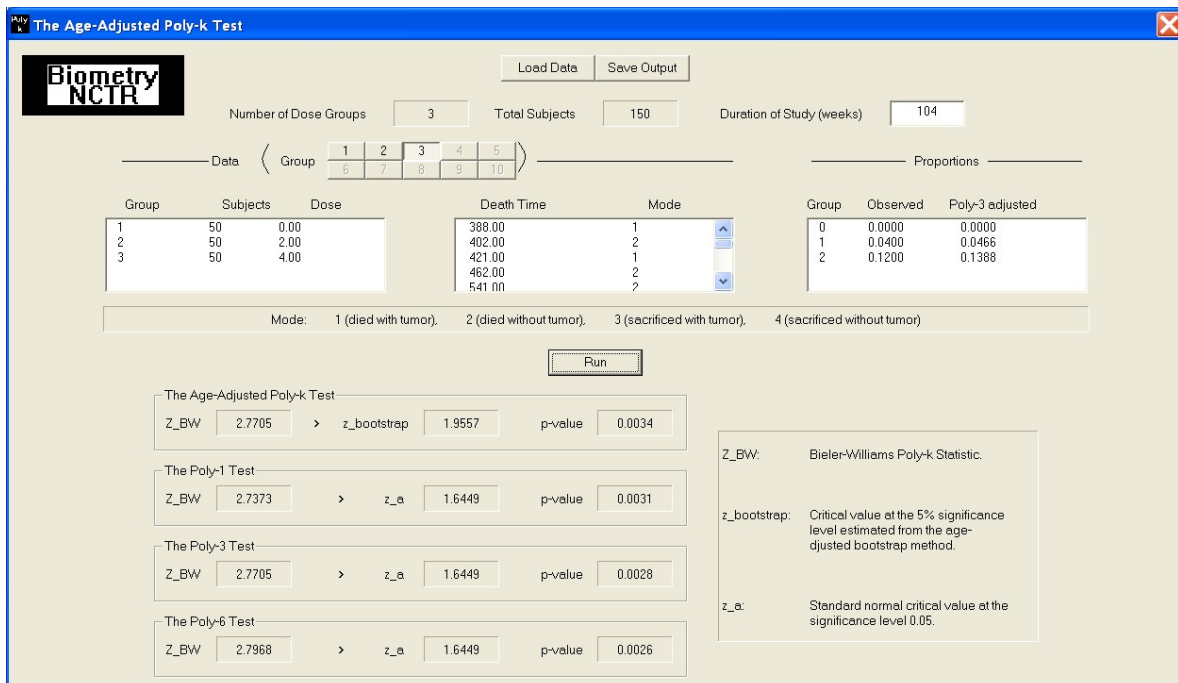

Figure 7: The age-adjusted poly-k test, the poly-1 test, the poly-3 test and the poly-6 test on the carcinogenicity activity of furan among three groups

statistical analysis of tumorigenicity data from animal bioassays via our age-adjusted poly-k test and the poly-k test.

References

Ahn H, Kodell RL (1998). *Analysis of Long-Term Carcinogenicity Studies.* Marcel Dekker, New York.

Armitage P (1955). "Tests for Linear Trends in Proportions and Frequencies." *Biometrics*, **11**, 375–386.

Armitage P (ed.) (1971). *Statistical Methods in Medical Research.* John Wiley, New York.

Bailer AJ, Portier CJ (1988). "Effects of Treatment-induced Mortality and Tumor-induced Mortality on Tests for Carcinogenicity in Small Samples." *Biometrics*, **44**, 417–431.

Bieler GS, Williams RL (1993). "Ratio Estimates, the Delta Method, and Quantal Response Tests for Increased Carcinogenicity." *Biometrics*, **49**, 793–801.

Cochran WG (1954). "Some Methods for Strengthening the Common χ^2 tests." *Biometrics*, **10**, 417–451.

Efron B, Tibshirani RJ (1993). *An Introduction to the Bootstrap.* Chapman and Hall, New York.

Farrar DB, Crump KS (1990). "Exact Statistical Tests for any Carcinogenic Effect in Animal Bioassays: II. Age-adjusted Tests." *Fundamental and Applied Toxicology*, **15**, 710–721.

Hoel D, Walburg HE (1972). "Statistical Analysis of Survival Experiments." *Journal of National Cancer Institute*, **49**, 361–372.

Kodell RL, Chen JJ, Moore GE (1994). "Comparing Distributions of Time to Onset of Disease in Animal Tumorigenicity Experiments." *Communications in Statistics – Theory and Methods*, **23**, 959–980.

Kodell RL, Lin KK, Thorn BT, Chen JJ (2000). "Bioassays of Shortened Duration for Drugs: Statistical Implications." *Toxicological Sciences*, **55**, 415–432.

Lee PN, Fry JS, Fairweather WR, Haseman JK, Kodell RL, Chen JJ, Roth AJ, Soper K, Morton D (2002). "Current Issues: Statistical Methods for Carcinogenicity Studies." *Toxicological Pathology*, **30**, 403–414.

Mancuso JY, Ahn H, Chen JJ, Mancuso JP (2002). "Age-adjusted Exact Trend Tests in the Event of Rare Occurrences." *Biometrics*, **58**, 403–412.

Mantel N, Haenszel W (1959). "Statistical Aspects of the Analysis of Data from Retrospective Studies of Disease." *Journal of National Cancer Institute*, **22**, 719–728.

Moon H, Ahn H, Kodell RL (2002). "Extension of Peto's Test by Attribution of Tumor Lethality in the Absence of Cause-of-Death Information." *Biometrical Journal*, **44**, 982–1001.

Moon H, Ahn H, Kodell RL (2005). "An Age-adjusted Bootstrap-based Poly-k Test." *Statistics in Medicine*, **24**, 1233–1244.

Moon H, Ahn H, Kodell RL, Lee JJ (2003). "Estimation of k for the Poly-k Test." *Statistics in Medicine*, **22**, 2619–2636.

National Toxicology Program (1993). "Toxicology and Carcinogenesis Studies of Furan in F344/N Rats and B6C3F$_1$ Mice (Gavage Studies)." *NTP Technical Report 402*, National Toxicology Program, Research Triangle Park, NC.

Office of the Federal Register (1985). "Chemical Carcinogens; A Review of the Science and Its Associated Principles." *Federal Register II*, Office of Science and Technology Policy.

Peto R, Pike MC, Day NE, Gray RG, Lee PN, Parish S, Peto J, Richards S, 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*, International Agency for Research on Cancer, Lyon, France.

STP Peto Analysis Working Group (2001). "The Society of Toxicologic Pathology's Position on Statistical Methods for Rodent Carcinogenicity Studies." *Toxicological Pathology*, **29**, 670–672.

STP Peto Analysis Working Group (2002). "The Society of Toxicologic Pathology's Recommendations on Statistical Analysis of Rodent Carcinogenicity Studies." *Toxicological Pathology*, **30**, 415–418.

US Food and Drug Administration (2001). "Guidance for Industry: Statistical Aspects of the Design, Analysis, and Interpretation of Chronic Rodent Carcinogenicity Studies of Pharmaceuticals." *Federal Register*, **66**, 23266–23267.

Woodruff RS (1971). "A Simple Method for Approximating the Variance of a Complicated Estimate." *Journal of the American Statistical Association*, **66**, 411–414.

Affiliation:

Hojin Moon
National Center for Toxicological Research
US Food and Drug Administration
3900 NCTR Road, HFT-20
Jefferson, AR 72079, United States of America
E-mail: hojin.moon@fda.hhs.gov