Recent Results Concerning Multiplicities in Animal Carcinogenicity Studies

Mohammad A. Rahman, Ph.D. Senior Mathematical Statistician,
and Karl K. Lin, Ph.D., Expert Mathematical Statistician
Division of Biometrics-6, FDA.

Disclaimer: The views expressed in this work are of the authors and not necessarily of the U.S. Food and Drug Administration

Presented in the BASS XIII
meeting in Savannah, Georgia,
November 6-10, 2006

Outline

- Brief description of carcinogenicity study design and data analysis.
- Analysis methods suggested by Peto et. al. and some concerns about this method.
- Poly-K method, an alternative method to Peto method.
- Issue of multiplicity of hypothesis testing in carcinogenicity data analysis.

Outline

- Methods of adjustments for multiplicity.
- An ad hoc method suggested by Lin and Rahman for Peto analysis.
- Examination of the adjustment method suggested by Lin and Rahman for Poly-K analysis.
- Conclusion.

Carcinogenicity Study Design

- Standard carcinogenicity study features
  - Two species: mouse and rat
  - Two sexes: male and female
  - Four groups: Control, low, medium, high
  - Group size: 50-70
  - Study duration: 2 years
  - Costs: between 1 and 2.5 million dollars

Carcinogenicity Study Design

- All animals alive at the end of 2 years are terminally sacrificed.
- All organs of all animals died during the study period or terminally sacrificed are microscopically examined for the presence of neoplastic and non-neoplastic lesions.

Alternative Model for Carcinogenicity Study

- The International Conference on Harmonization (ICH) has recommended to use either two 2-year studies in rats and mice or alternatively one 2-year study in rats along with one short or medium-term study in transgenic mice that provide rapid observation of carcinogenic endpoints.

Transgenic Strains

The main transgenic strains proposed and used include:

- p53+/- transgenic mice (with knockout of one of the two alleles of the tumor suppression gene p53)
- Tg.AC transgenic mice (with genetically initiated skin to induce epidermal papillomas in response to dermal or oral exposure to chemical agents and act as a reporter phenotype of the activities of the tested chemicals)
- Tg rasH2 transgenic mice (with 5-6 copies of the stable human c-Ha-ras gene, developed and patented in Japan).
- XPA+/- repair deficient transgenic mice (developed in Europe).

Transgenic Mouse Study Designs

- For studies with Tg rasH2, p53+/-, and XPA+/-, there will be a positive control group (treated with p-cresidine, or benzene, or TPA) in addition to the regular 3 or 4 treatment groups (negative control, low, medium, and high)
- 15-25 animals per sex/treatment group.
- 26 weeks of duration
- Like the long term studies, all Tissues/Organs of p53+/-, Tg rasH2, and XPA+/- repair deficient mice died or terminally sacrificed are microscopically examined for neoplastic and non-neoplastic lesions.

Transgenic Mouse Study Designs

- In Tg.AC transgenic mice studies, only weekly incidence rates and weekly counts of skin papillomas are collected for evaluation of drug effects.

Analysis of 2-Year Carcinogenicity Study Data

Analysis of 2-year carcinogenicity data includes:

- Test for dose response (positive trend) among the increasing doses by organ/tumor combination.
- Pairwise comparisons of treated groups with control also by organ/tumor combination.

Analysis of 26-Week Transgenic Mouse Study Data

- To analyze data of p53+/-, Tg.rasH2, and XPA+/- mice studies, statisticians in CDER use similar analyses as used for the 2-year studies i.e. test for dose response and pairwise comparisons

Analysis of 26-Week Transgenic Mouse Study Data

- Methods of analyzing data from the Tg.AC transgenic mice study are different from other models because of the difference of collected endpoints.

Analysis of 26-Week Transgenic Mouse Study Data

- Dunson et al.,(2000) developed a method to analyze this data. Their method separates the drug effects on papillomas into latency and multiplicity, and accommodates important features of the data, including variability in expression of the transgene and dependency in the tumor counts.

Important Adjustments Needed in Carcinogenicity Data Analysis

- Adjustment for differences in mortalities among groups.
  -Animals living longer have high probability of developing tumors. This is particularly important for 2-year studies. For 26-week study low mortality makes the survival adjustment less important.

Important Adjustments Needed in Carcinogenicity Data Analysis

- Multiplicity adjustment.
  -Because the analysis for positive trend is done by organ/tumor combination, a large number of comparisons are involved. Therefore, a great potential exists for finding statistically significant dose response or treatment-placebo differences due to chance alone (i.e., a false positive).

Statistical Methods for Data Analysis

- There are many methods suggested for carcinogenicity data analysis. However, because of the relevance, in this presentation we will discuss only three methods namely, Cochran-Armitage test, test proposed by Peto et al. (Peto test), and Poly-K test.

Cochran-Armitage Test

- If there is no significant survival difference among dose groups Cochran-Armitage test may be used on the proportions of animals with tumor.

Suppose, $\hat{p}_i = \frac{x_i}{n_i}$, $I = 0, \ldots, r$

Be the observed proportion of tumor bearing animals in i^{th} group, where x_i= number of animals with tumor in the i^{th} dose group and n_i= total number of animal in the i^{th} dose group.

Cochran-Armitage Test

- To test the dose-response in p_i in r+1 dose groups, Cochran and Armitage suggested the statistic

$$\chi^2 = (\sum x_i d_i - \hat{p} \sum n_i d_i)^2 / \{\hat{p}\,\hat{q}\,[\sum n_i d_i^2 - (\sum x_i d_i)^2 / n.]\}$$

where, $n. = \sum n_i$, $x. = \sum x_i$, $\hat{p} = x. / n.$, and $\hat{q} = 1 - \hat{p}$.

- The test statistic χ^2 is distributed asymptotically as a Chi square with one degree of freedom.

Peto Test

- As mentioned before, if there is significant survival difference among dose groups, analysis method needs adjustment for mortality.
- Method suggested by Peto et al. adjusts the mortality differences by dividing the study period into several intervals and using the cause of death information. The method analyzes data in each interval and combines results using the Mantel-Haenszel procedure.

Peto Test

- In Peto's method the denominator and numerator for the calculation of proportion of tumor bearing animals is determined based on the cause of death information.
- If tumor caused the death then the number of animals alive at the time of detection is the denominator and the number of animals detected with the tumor at the time of detection is the numerator.
- If tumor did not cause the death then the number of animals dying in the entire time interval of detection is the denominator and the number of animals detected with the tumor in the entire time interval is the numerator.

Cause of Death Information

- According to the opinion of many pathologists it is very difficult to correctly specify if the related tumor is the real cause of death of an animal. This information may be imprecise.
- Hence, many times results of analysis using methods suggested by Peto et al. are questioned.

Poly-K Test

Blair and Portier (1988) suggested an alternative method, known as the Poly-K test, to adjust for mortality differences which does not need the cause of death information. This method considers less than a whole animal for animals dying early without tumor.

Poly-K Test

The proportion of animals with tumor is calculated as,

$$p_i^* = \frac{x_i}{n_i^*}$$

Where $n_i^* = \Sigma w_{ij}$, and

$w_i=(t_{ij}/t_{max})^k$ for animals dying before the end of the study without tumor, and 1 otherwise.

t_{ij}= Survival time of j^{th} animal in i^{th} group.

Value of K depend on the mortality pattern of the animals. For carcinogenicity data analysis K=3 is mostly in use.

Poly-K Test

- These modified proportions are then analyzed using the Cochran-Armitage test. However, since n* is a random variable (not fixed as required by the Cochran-Armitage test) calculation of variance of test statistic needs to be modified.
- Bieler and Williams (1993) suggested an estimate of this variance, using delta method and weighted least squares technique.

Poly-K Test

- P-values calculated by the Poly-k method may differ, some times markedly, from those calculated using the Peto method.

Example of Peto and Poly-k P-Values from a 2-Year Study with Significant Survival Difference

		Poly-k		
	Peto	K=1	K=3	K=6
Adrenal/Adenoma	0.148	0.080	0.085	0.089
Adrenal/Medullary	0.162	0.069	0.073	0.078
Lungs/Carcinoma	0.043	0.015	0.018	0.021
Skin/Fibroma	0.149	0.069	0.074	0.079
Uterus/Leiomyoma	0.425	0.303	0.315	0.326
Uterus/Granullar cell	0.169	0.131	0.143	0.155
Stomach/Adenocar	0.271	0.074	0.070	0.069
Liver/Hemangioma	0.639	0.527	0.558	0.579
Uterus/Adenoma	1.000	0.918	0.916	0.916

Multiplicity Issue

- Since the analysis is performed by tumor/organ combinations, and there are usually 2 species, 2 sexes, and 30 or more tumor/organ examined, a total of 120 or more tumor/organ types are tested in total 2 year studies. Hence, the issue of multiplicity is severe.
- The 26-week studies generally show a smaller number of tumor types. Hence, adjustment for multiplicity may be less sever for 26-week studies compared to 2-year studies.
- It should be noted that the multiplicity issue is serious for common tumors than rare tumors.

Multiplicity Issue

- Operationally we want to control the overall false positive error rate over the four experiments (one for each sex/species).
- There are many available methods for this purpose e.g. methods suggested by Bonferroni, Hochberg, false discovery rate, method based on bootstrap/re-sampling, and method based on Bayesian approach. Also there are some ad hoc methods developed specially for carcinogenicity data analysis.

Multiplicity Issue

- One ad hoc method for dose response tests suggested by Lin and Rahman (1992) is the use of α=0.025 for rare tumor and α=0.005 for common tumor.
- Rare tumor type is defined as tumor with less than 1% background rate.

Multiplicity Issue

- Lin and Rahman showed through some empirical and simulation studies with 20 organ/tumor types in each sex/species with various background rates that use of their recommended levels of significance on dose response tests using Peto method resulted in an overall false positive rate of about 10% .

FDA Empirical Study

- A spontaneous tumor rate data of Crt:CDBR rats and Crt:CD-1(ICR) BR mice were compiled in the Division of Biometrics using information provided by the Charles River Company. In their empirical study Lin and Rahman used prevalence rates from this data set.

FDA Empirical Study

- The following assumptions were made

  1. There were 4 treatment groups -3 treated at equally spaced doses and one control

  2. There were 50 animals in each treatment/sex group

  3. Tumor occurred independently of each other

FDA Empirical Study

- The overall false positive rates in different data sets were between 8% to 13% with an average of about 10%.
- As mentioned in previous slide, one of the assumptions was independence of tumor types. However, often some positive dependence is found among tumor types. As an effect of such dependence, the actual overall false positive rate may be even less that 10%.

Multiplicity Issue

- The overall false positive rate using of Lin and Rahman's method in Poly-K analysis in not known.
- To examine this overall false positive rate we performed a simulation study.

Simulation to Compare Peto and Poly-K Methods in 2 Year Study

- In this simulation, we first computed the overall false positive rates using the Peto and Poly-K test from the same simulated data set.
- Conclusions on the performance of Lin and Rahman's adjustment method in the Poly-K test was then drawn based on the closeness of the overall false positive rates from the two methods.

Simulation to Compare Peto and Poly-K Methods in 2 Year Study

- Results from 26-week study were investigated later through similar simulation.

Simulation to Compare Peto and Poly-K Methods in 2 Year Study

- Tumor data were generated for 4 treatment groups with increasing doses (Control, Low, Medium, and High dose group).
- There were 50 animals per group.
- Study length were 2 years (104 weeks). All animals survived after that were considered as terminally sacrificed.
- All generated tumors were assumed to be incidental.

Simulation to Compare Peto and Poly-K Methods in 2 Year Study

- Tumor detection time and survival time were modeled by four parameters Weibull distributions.
- Survival time was modeled as

  $S(t/x)= P(t>t/x)\exp\{-(C+Dx)(t-A)^B$, and
- Tumor detection time was modeled as $P(t,x)=1 – S(t/x)$.

Simulation to Compare Peto and Poly-K Methods in 2 Year Study

- Two vectors, each with 200x1 dimension were generated from two Weibull distributions representing the time of death and tumor onset time of 200 animals.
- For an animal if tumor onset time was less than or equal to time of death the animal was assumed to develop the tumor.
- Actual tumor detection time was assumed to be the time of death.

Simulation to Compare Peto and Poly-K Methods in 2 Year Study

- Two hundred animals were then randomly allocated to four treatment groups of 50 animals each.
- Each animal was assumed to be equally likely to develop tumors in their life time.
- Tumors were develop independent of each other.

Simulation to Compare Peto and Poly-K Methods in 2 Year Study

- P-values were calculated using both Peto and Poly-K methods.
- Re-allocation and p-value calculation were repeated for 1000 times.
- Number of significant results was counted and false positive rate was calculated.

Simulation to Compare Peto and Poly-K Methods in 2 Year Study

- For Peto analysis, the survival time was grouped into five intervals namely, 0-52, 53-78, 79-92, 93-104 weeks, and terminal sacrifice. Scores used were 0, 1, 2, and 3 for control, low, medium, and high dose group.
- For Poly-K method, K=3 was used.

Model Parameters

Model	Time to Event	Weibull Parameters A	B	C x 10^4	D x10^4	Effect of drug on death
1	Time to death (T1)	0	4	0.0000305	0	NO
2	Time to death (T1)	0	4	0.0000305	0.0000239	SMALL
3	Time to death (T1)	0	4	0.0000305	0.00008325	LARGE
For each model for time to death						
(a)	Time to tumor (T0)	17	2	0.0005	0	
(b)	Time to tumor (T0)	17	2	0.0015	0	
(c)	Time to tumor (T0)	17	2	0.0025	0	
(d)	Time to tumor (T0)	17	2	0.0050	0	
(e)	Time to tumor (T0)	17	2	0.0080	0	
(f)	Time to tumor (T0)	17	2	0.01000	0	
(g)	Time to tumor (T0)	17	2	0.01500	0	
(h)	Time to tumor (T0)	17	2	0.02000	0	

Model Parameters

- Since, within each model, all parameters are constant except ‘c’ and the probability of an event is a direct function of ‘c’ (defined as the baseline scale parameter), results will be interpreted in terms of ‘c’.

Results

Method	Effect of Drug on Death	Weibull Parameter c	Peto Method P-Values with level 0.05	Peto Method P-Values with level 0.025 and 0.005	Poly-3 P-Values with level 0.05	Poly-3 P-Values with level 0.025 and 0.005
NO		0.0005	0.000	0.013	0.000	0.002
NO		0.0015	0.000	0.022	0.000	0.027
NO		0.0025	0.046	0.000	0.048	0.000
NO		0.0050	0.055	0.017	0.054	0.017
NO		0.0080	0.054	0.011	0.052	0.012
NO		0.0100	0.043	0.005	0.043	0.010
NO		0.0150	0.063	0.010	0.062	0.011
NO		0.0200	0.052	0.005	0.053	0.008

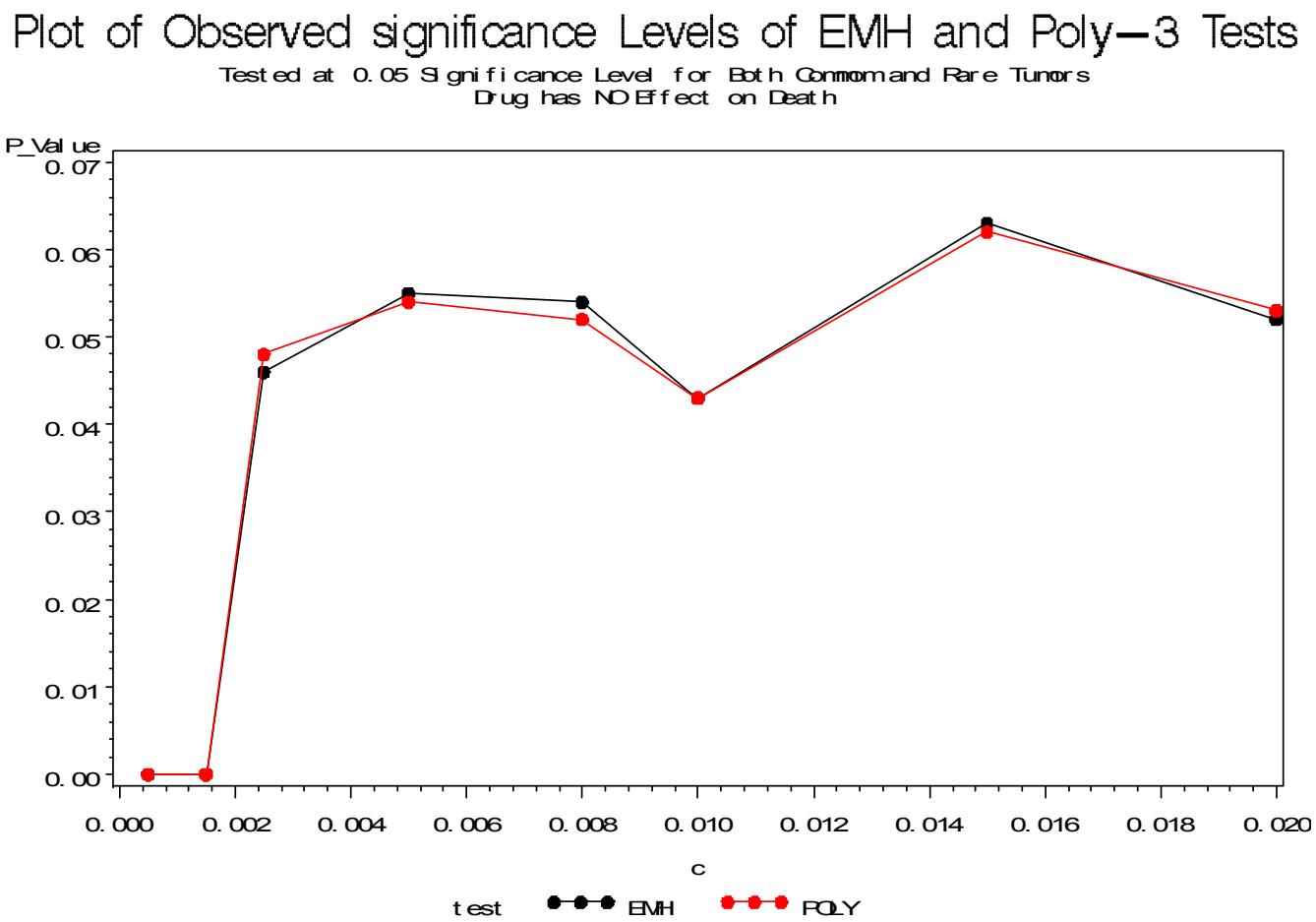
Results

Effect of Drug on Death	Weibull Parameter c	Peto Method P-Values with level 0.05	Peto Method P-Values with level 0.025 and 0.005	Poly-3 Method P-Values with level 0.05	Poly-3 Method P-Values with level 0.025 and 0.005
SMALL	0.0005	0.023	0.002	0.000	0.000
SMALL	0.0015	0.054	0.026	0.054	0.005
SMALL	0.0025	0.050	0.020	0.052	0.017
SMALL	0.0050	0.056	0.020	0.059	0.020
SMALL	0.0080	0.051	0.012	0.053	0.014
SMALL	0.0100	0.060	0.005	0.061	0.005
SMALL	0.0150	0.055	0.012	0.055	0.011
SMALL	0.0200	0.053	0.007	0.060	0.007

Results

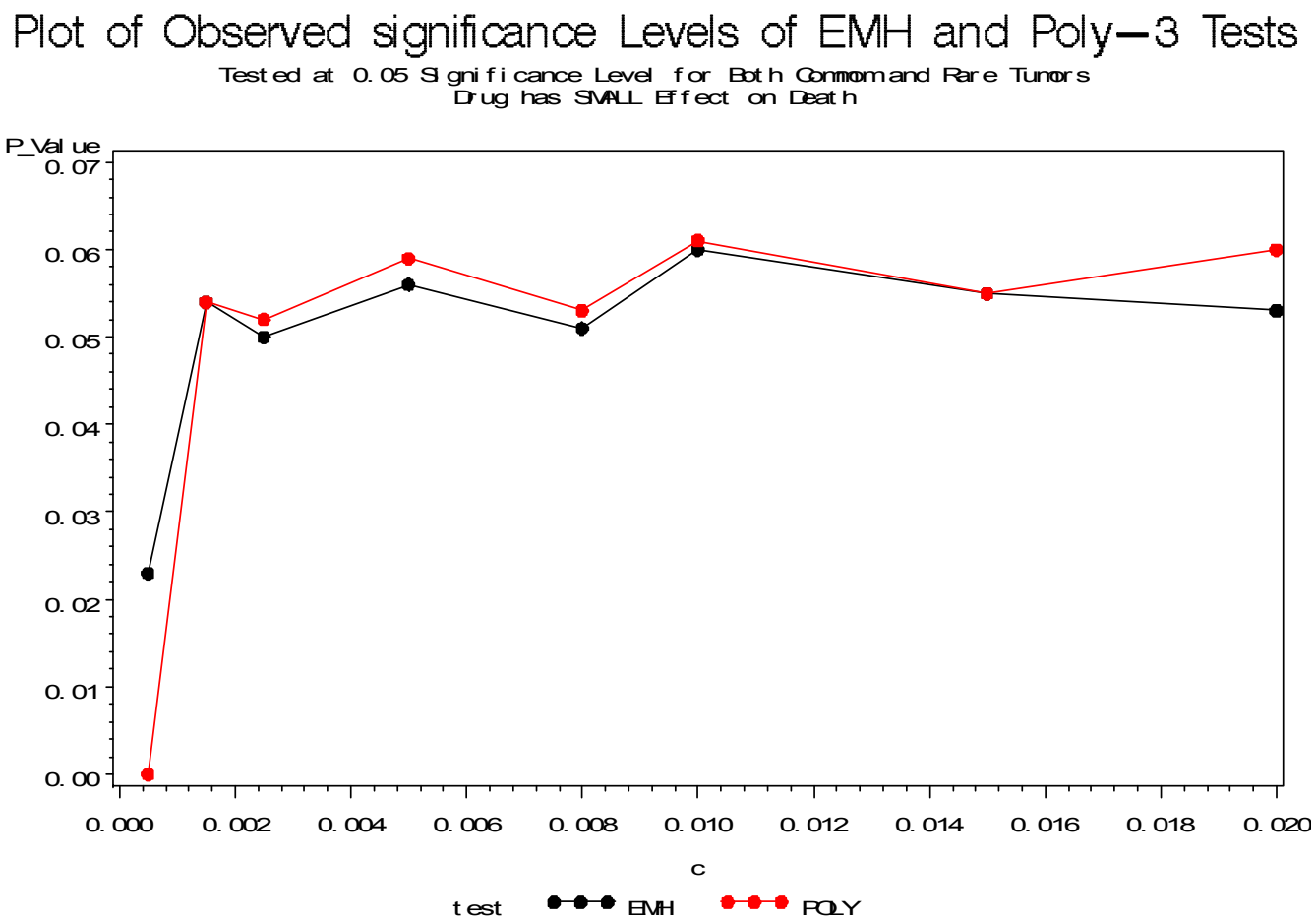
Effect of Drug on Death	Weibull Parameter c	Peto Method P-Values with level 0.05	Peto Method P-Values with level 0.025 and 0.005	Poly-3 Method P-Values with level 0.05	Poly-3 Method P-Values with level 0.025 and 0.005
LARGE	0.0005	0.005	0.001	0.000	0.000
LARGE	0.0015	0.054	0.022	0.055	0.023
LARGE	0.0025	0.060	0.020	0.066	0.023
LARGE	0.0050	0.044	0.031	0.044	0.028
LARGE	0.0080	0.055	0.013	0.054	0.014
LARGE	0.0100	0.053	0.010	0.053	0.012
LARGE	0.0150	0.040	0.008	0.044	0.008
LARGE	0.0200	0.045	0.005	0.048	0.006

Results



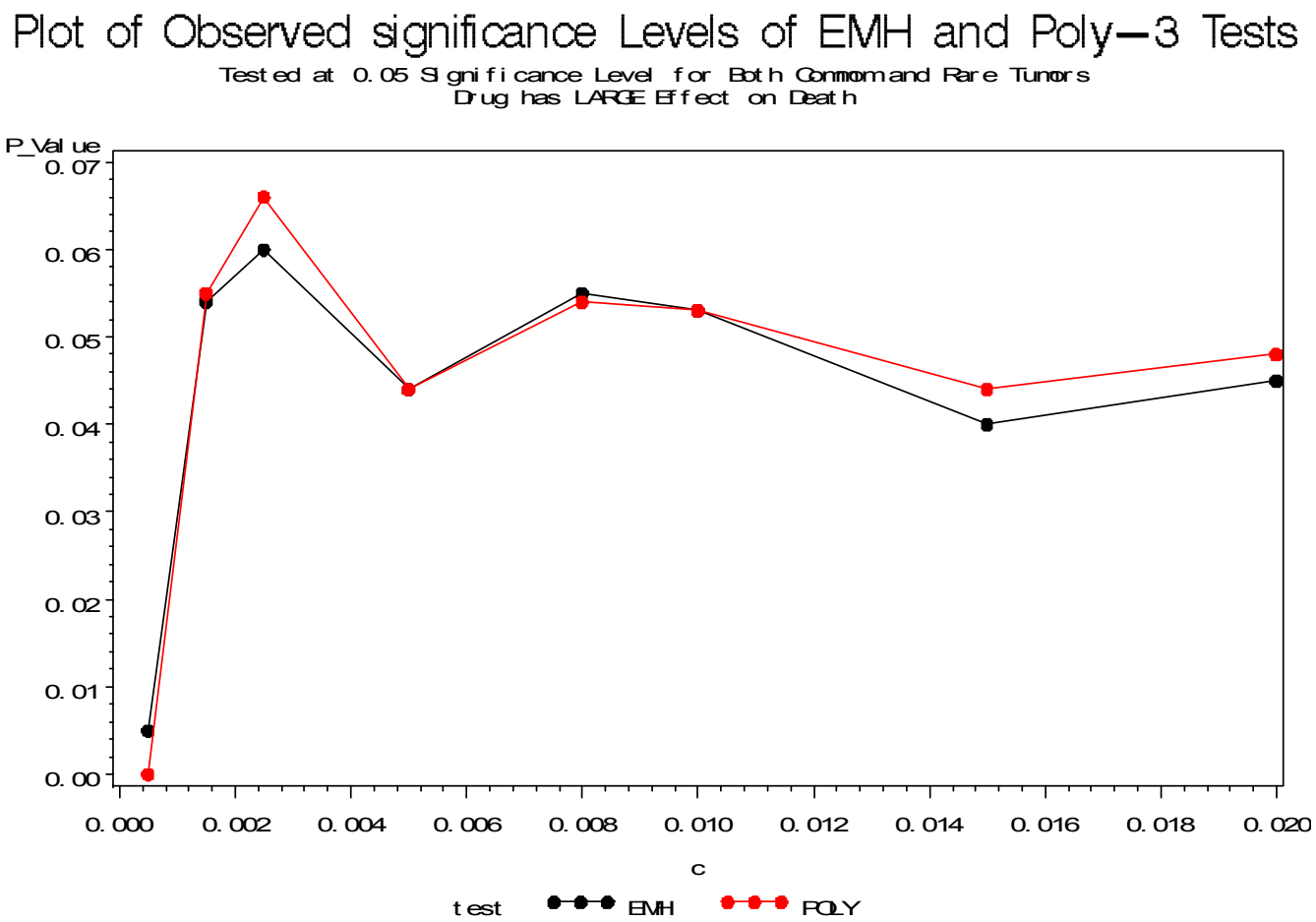
Presented in the BASS XIII meeting in Savannah, Georgia, November 6-10, 2006

Results



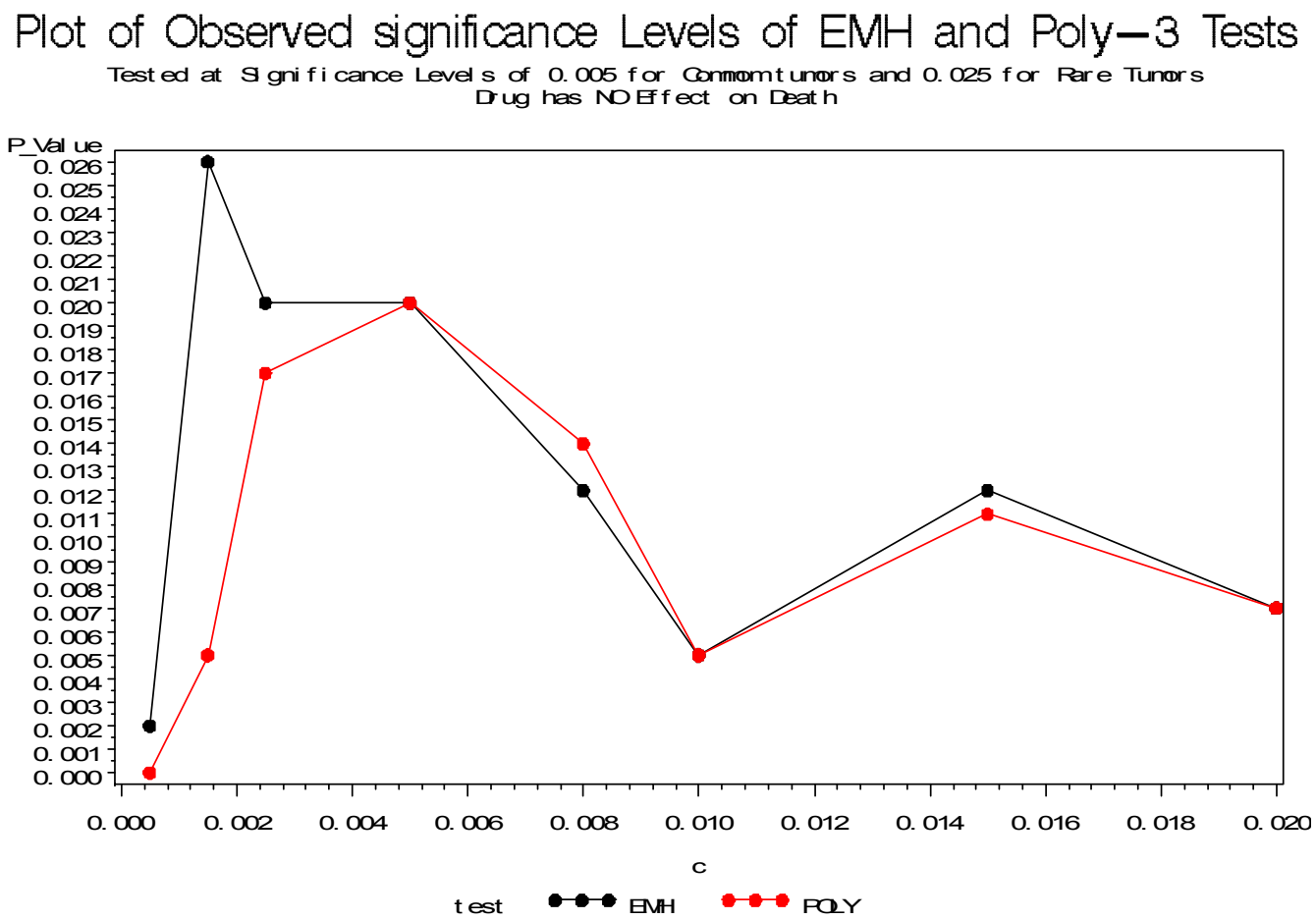
Presented in the BASS XIII meeting in Savannah, Georgia, November 6-10, 2006

Results



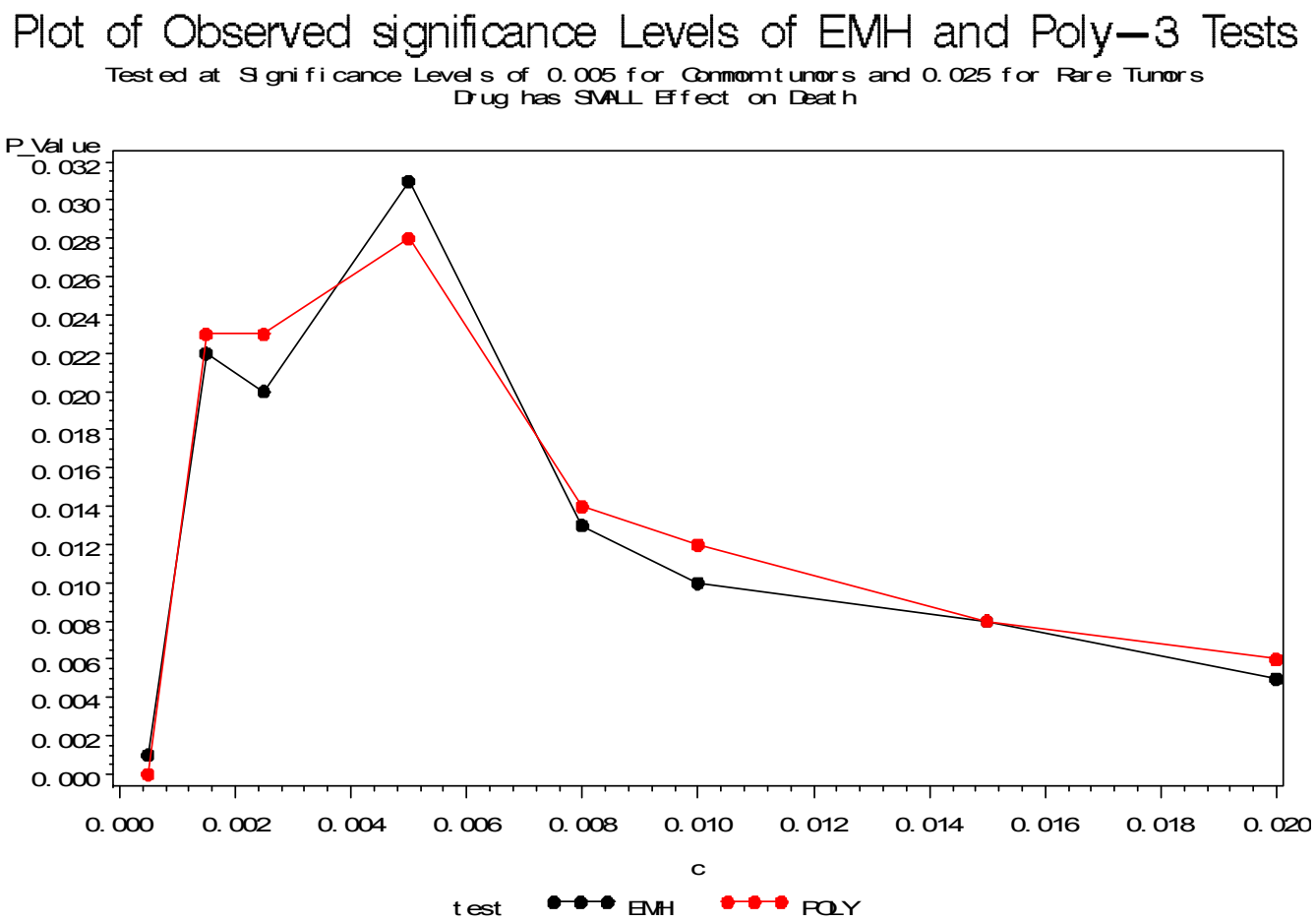
Presented in the BASS XIII
meeting in Savannah, Georgia,
November 6-10, 2006

Results



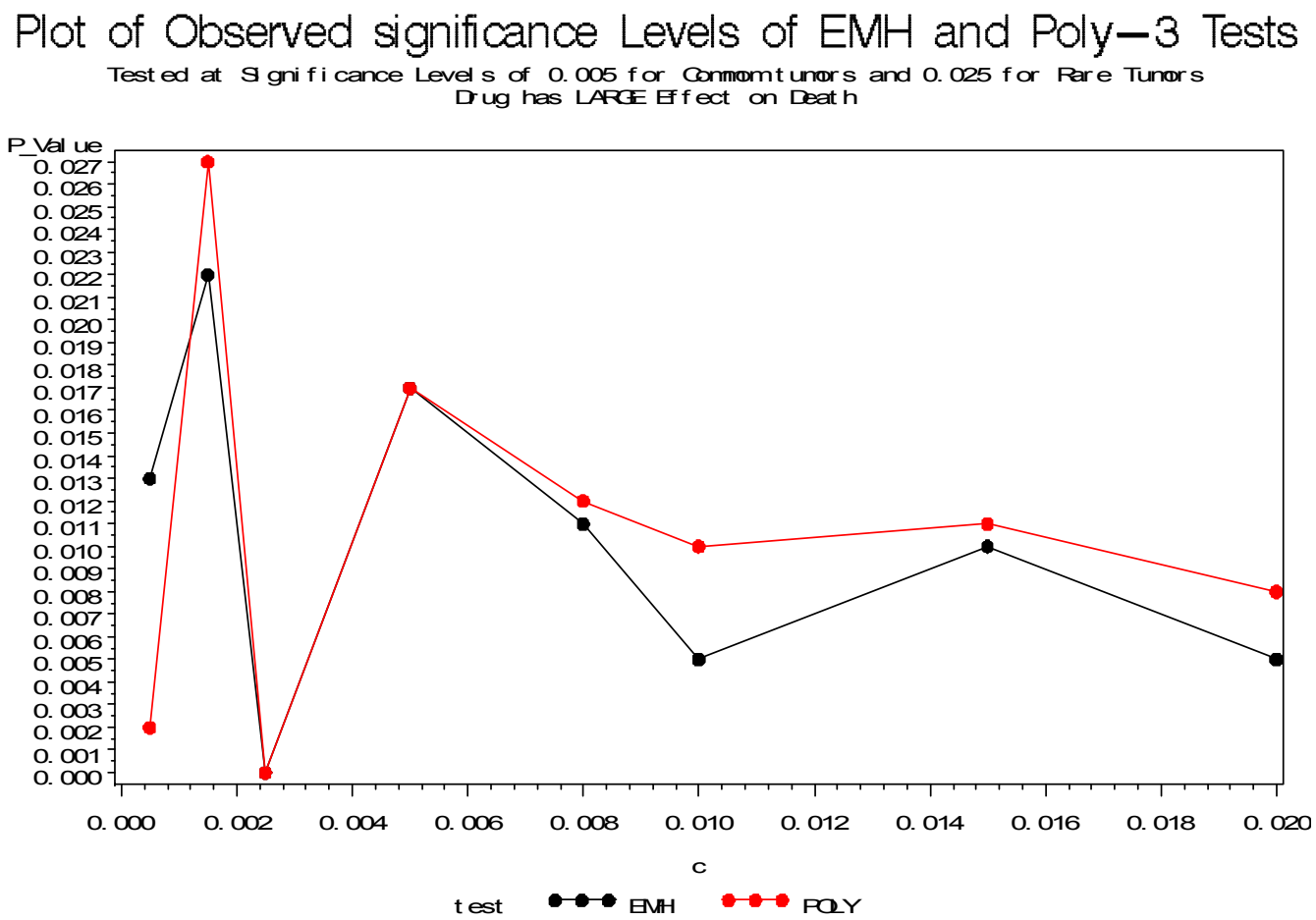
Presented in the BASS XIII
meeting in Savannah, Georgia,
November 6-10, 2006

Results



Presented in the BASS XIII meeting in Savannah, Georgia, November 6-10, 2006

Results



Presented in the BASS XIII
meeting in Savannah, Georgia,
November 6-10, 2006

Twenty Six Week Study in Transgenic Mouse

- Not enough data from 26 week transgenic mouse study is available in the division at this moment.
- Based in the limited experience, the data showed the occurrence of about 10/15 observed tumor types in each study, most with very low occurrence.
- Therefore, for transgenic mouse study very little or no adjustment might be needed.

Twenty Six Week Study in Transgenic Mouse

- At this moment, in the division a test level of $\alpha=0.05$ is used for all tumor types. However, with all the limitations of data and experience, some preliminary study showed that a test level of $\alpha=0.025$ for common tumor and $\alpha=0.05$ for rare tumor might be more suitable.
- More investigation in is needed this area as the data becomes available.

Conclusions

- The simulation and the empirical results shows that for 2-yaer regular rodent carcinogenicity study, the false positive rates using the Peto method and Poly-3 method are very close.
- It seems that for 2-yaer regular rodent carcinogenicity study the Lin and Rahman method is also applicable to Poly-3 analysis of two years animal carcinogenicity data.

Conclusions

- Preliminary results showed that, for 26 week transgenic mouse study, a test level of $\alpha=0.025$ for common tumor and $\alpha=0.05$ for rare tumor might be suitable. However, more investigation is needed to confirm this result.
- The authors, in collaboration with other colleagues, are investigating new methods e.g. method based on total number of significant findings, to handle high dimensional multiple testing problems.

Thanks

- The authors are grateful to Dr. Huque for his inspirations and comments on this work.

References

- Armitage, P. (1955), “Tests for Linear Trends in Proportions and Frequencies,” Biometrics, 11, 375-386.
- Peto, R., M.C. Pike, N.E. Day, R.G. Gray, P.N. Lee, S. Parish, J. Peto, S. Richards, and J. Wahrendorf (1980), "Guidelines for Simple, Sensitive Significance Tests for Carcinogenic Effects in Long-term Animal Experiments," in Long-term and Short-term Screening Assays for Carcinogens: An Critical Appraisal, World Health Organization.
- ICH (1998), S1B Testing for Carcinogenicity of Pharmaceuticals, ICH - S1B, Federal Register, vol. 63, 8983-8986, 1998.
- Bailer, A., and C. Portier (1988), “Effects of Treatment-Induced Mortality on Tests for Carcinogenicity in Small Samples,” Biometrics, 44, 417-431.

References

- Bieler, G.S., and R.L. Williams (1993), “Ratio Estimates, the Delta Method, and Quantal Response Tests for Increased Carcinogenicity,” Biometrics, 49, 793-801.
- Dunson, D.B., et al. (2000), “Statistical Analysis of skin Tumor data from Tg.AC Mouse Bioassays”, Toxicological Sciences, 55, 293-302.
- Lin K. Karl and Rahman A. Mohammad: Overall False Positive Rates in Tests for Linear Trend in Tumor incidence in Animal Carcinogenicity Studies of New Drugs. Journal of Biopharmaceutical Statistics, 8(1), 1-15 ,1998.