Department of Pesticide Regulation

Brian R. Leahy
Director

Edmund G. Brown Jr.
Governor

MEMORANDUM

TO: Sheryl Beauvais
Senior Toxicologist
Worker Health and Safety Branch

HSM-13001

FROM: Holly Jessop *[Original signed by H. Jessop]*
Research Scientist I (Epidemiology/Biostatistics)
916-324-3517

DATE: April 3, 2013

SUBJECT: RELEVANCE OF THE BIELER-WILLIAMS POLY-3 TREND TEST IN THE ANALYSIS OF CARCINOGENICITY OF CHLOROPICRIN

The Poly-3 Trend Test is a statistical procedure for the analysis of experimental data on carcinogenicity in animals that are exposed to a particular chemical compound. The data are analyzed to assess whether there is a relationship between the amount of compound administered and the presence of tumors after sacrifice; carcinogenesis as a function of compound treatment level is the outcome of interest. For typical experiments in which animals are sacrificed at a pre-set time and the cause of any early death is unknown, analysis challenges arise regarding the possibility of 'tumor lethality' (which occurs when an animal dies due to the development of cancer before the pre-set sacrifice date) and 'treatment lethality' (which occurs when animals in groups with a higher dose of compound are less likely to live long enough for a tumor to develop). Both tumor and treatment lethality, which can influence survival time, can easily bias the results of an analysis for carcinogenicity (Bailer and Portier, 1988a; Bailer and Portier, 1988b; Bieler and Williams, 1993; Peto et al., 1980). Bailer and Portier (1988a) introduced the Poly-3 Test to take such survival time differences into account, thereby improving an existing statistical procedure for the analysis of carcinogenicity data that inaccurately assumes that all animals have equal risk for developing a tumor over the course of the study period (i.e., the Cochran-Armitage Linear Trend Test). However, especially for low incidence tumors, the Poly-3 Test can tend to incorrectly reject the Null Hypothesis (i.e., to incorrectly conclude that a compound causes cancer in experimental animals, known as a Type I Error), due to overestimating the tumor incidence rates in higher dose groups (Bailer and Portier, 1988a; Bieler and Williams, 1993). Later, Bieler and Williams (1993) made additional improvements to the Poly-3 Test that ameliorate this problem; the Bieler-Williams Poly-3 Test has been shown to be robust (i.e., to yield an unbiased statistical result), regardless of the level of tumor or treatment lethality, including for low background tumor rates (Bieler and Williams, 1993).

A carcinogenicity study of the compound chloropicrin entitled Project 92-N1106 involved 50 CD-1 mice per sex per dose (0, 0.1, 0.5, or 1.0 ppm), with a pre-set sacrifice time of 82 weeks (Burleigh-Flayer et al., 1994). No treatment-related effects on survival were observed, and so these tumor data were initially analyzed by Lewis (2010) using a Fisher's Exact Test, finding a p-value of 0.053 for a tumor (lung adenoma and carcinoma) increase in the high dose female mice. Such a result ($p > 0.05$) is often interpreted as statistically non-significant; the

interpretation of the p-value is that if chloropicrin does *not* cause cancer in mice, then the probability that one would observe such an incidence of tumors in the high dose female mice by chance alone is 0.053 (or, slightly more than about 5 times in 100). Lewis (2010) also analyzed these same data using the Bieler-Williams Poly-3 Test, finding a p-value of 0.03. Such a result ($p < 0.05$) is often interpreted as statistically significant; again, the interpretation of the p-value is that if chloropicrin does *not* cause cancer in mice, then the probability that one would observe such an incidence of tumors by chance alone is 0.03. Such a probability, of less than about 5 times in 100, is typically considered unlikely.

The usage of the Bieler-Williams Poly-3 Test by Lewis in the analysis of the Project 92-N1106 study can be criticized in a number of ways, including: 1) Survival times did not differ between treatment groups in this particular study, which may negate the need for such a survival-adjusted statistical approach. However, as Lewis (2010) points out, the seminal carcinogenic statistical methods paper by Peto et al. (1980) recommends that tumor rates should *always be adjusted for survival*, even when there has not been an observed difference in survival times. Furthermore, the National Toxicology Program (NTP) has also adopted such an analysis policy; 2) The ability to extrapolate the Bieler-Williams Poly-3 Test from the B6C3F1 mouse strain (with which the test was validated by data from the NTP (Bailer and Portier, 1988a; Portier et al., 1986)), to the CD-1 strain of mice utilized in this particular study, is untested. However, as Lewis (2010) mentions, historical control data for lung adenomas in female CD-1 mice (0-27%) are similar to that of female B6C3F1 mice (0-24%), which provides some evidence that the tumor onset rate, and thereby the mathematical model utilized by the Bieler-Williams Poly-3 Test, is not significantly different between these two mouse strains.

Whether or not one considers the statistical analysis of the Project 92-N1106 study to be more correct using a Fisher's Exact Test, or the Bieler-Williams Poly-3 Test -- *both methods yield equivocal results when applied to this particular study*. First, while a p-value significance level of 0.050 is a common boundary for setting statistical significance, it is also a somewhat arbitrary cut-off point. Second, a p-value cannot be taken as an exact literal probability value, and a p-value does *not* equate to the probability that the Null Hypothesis is correct. Thus, a p-value of slightly more than 0.05 is not sufficient reason to conclude that chloropicrin did not cause lung cancer in these experimental mice. And conversely, a p-value of slightly less than 0.05 is not reason enough to conclusively state that chloropicrin did cause lung cancer in this particular study. Instead, p-values should be utilized as general guidance for making causal inferences, within the context of an overall evaluation of the merits and detractions of the study, and together with considerations of other independent evidence.

In the case of the Project 92-N1106 study, the p-values from both statistical analysis approaches provide an indication that the data are not very consistent with the Null Hypothesis (i.e., the hypothesis that chloropicrin does not cause cancer in mice). So, the results from this particular study, regardless of the manner in which the data are statistically analyzed, provide some (albeit

borderline) evidence that high dose female mice develop increased numbers of lung tumors. However, while this study does provide some evidence that the carcinogenic strength of chloropicrin is likely not zero, neither does the study provide evidence that this compound is likely to be a strong carcinogen, given the borderline p-values.

References

Bailer, A.J., and Portier, C.J. 1988a. Effects of treatment-induced mortality and tumor-induced mortality on tests for carcinogenicity in small samples. Biometrics 44: 417-431.

Bailer, A.J., and Portier, C.J. 1988b. An illustration of dangers of ignoring survival differences in carcinogenic data. Journal of Applied Toxicology 8: 185-189.

Bieler, G.S., and Williams, R.L. 1993. Ratio estimates, the Delta Method, and quantal response tests for increased carcinogenicity. Biometrics 49: 793-801.

Burleigh-Flayer, H.D., Kintigh, W.J., Benson, C.L. 1994. Chloropicrin: Vapor inhalation oncogenicity study in CD-1 Mice (BRRC Project 92-N1106). Bushy Run Research Center, Export, PA.

Lewis, C.M. 2010. Evaluation of chloropicrin as a toxic air contaminant. Part B, Human Health Assessment. California Environmental Protection Agency, Department of Pesticide Regulation, Medical Toxicology Branch.

Peto, R., Pike, M., Day, N., Gray, R., Lee, P., Parish, S., Peto, J., Richard, S., and Wahrendorf, J. 1980. Guidelines for simple, sensitive significance tests for carcinogenic effects in long-term animal experiments. In Annex to Long-Term and Short-Term Screening Assays for Carcinogens: A Critical Appraisal, International Agency for Research on Cancer Monographs, Supplement 3, 331-426. Lyon: IARC.

Portier, C., Hedges, J., and Hoel D. 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.

cc: Marylou Verder-Carlos, Assistant Director
Lisa Ross, Chief, Worker Health and Safety Branch