FUNDAMENTAL AND APPLIED TOXICOLOGY 12, 731–737 (1989)

Testing for Increased Carcinogenicity Using a Survival-Adjusted Quantal Response Test

C. J. PORTIER AND A. J. BAILER

Statistics and Biomathematics Branch, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina 27709

Received May 27, 1988; accepted December 2, 1988

Testing For Increased Carcinogenicity Using A Survival-Adjusted Quantal Response Test. PORTIER, C. J., AND BAILER, A. J. (1989). *Fundam. Appl. Toxicol.* **12**, 731–737. The linear trend test in proportions is frequently used to analyze the results of animal carcinogenicity experiments. This test has two major advantages over other frequently used tests; it is easily understood and it is simple to calculate. This test, however, fails to correct for treatment-related differences in survival across the experimental groups. A test which is a simple modification of the linear trend test in proportions and which has the same advantages is proposed to correct for differences in survival. The results of this modified test are compared to those of the linear trend test in proportions, the incidental tumor test, the logistic regression score test, the life table test, and the truncated trend test using information on the incidence of combined alveolar/bronchiolar adenomas or carcinomas in female $B6C3F_1$ mice exposed to vinylcyclohexene diepoxide. © 1989 Society of Toxicology.

The statistical analysis of carcinogenicity data has evolved in the last few years from simple contingency table analyses to more complicated survival analytic procedures. Most of these methods are discussed in general terms in a recent review by McKnight (1988). The result of this evolution is that toxicologists and biostatisticians presented with the task of analyzing an animal carcinogenicity experiment have a considerable number of methods from which to choose.

Hoel and Walburg (1972) recognized the need to correct for survival differences when analyzing carcinogenicity data. For occult tumors, the time at which tumor onset occurs is unobservable. What is observed is the animal's death time and the presence or absence of the tumor at death. Since the tumor can affect the age at which the animal dies, studying only animals that die naturally can lead to biased estimates of tumor incidence. To protect against this bias, the degree to which the presence of a tumor increases an animal's chances of dying must be estimated. Recently developed techniques appear to provide unbiased estimates of tumor incidence for some carcinogenicity experiments (e.g., McKnight and Crowley, 1984; Dewanji and Kalbfleisch, 1986; Portier, 1986; Portier and Dinse, 1987). However, these approaches have drawbacks which make them unusable in many situations. In the first place, these procedures require the use of interim terminations in the experiment. The cost of incorporating additional animals into the experiment can be substantial, making this modification unattractive. In addition, there are a substantial number of past experiments which have no interim terminations but which exhibit differential survival in the treated groups.

This paper presents a test of carcinogenicity which is simple, is not affected by the de-

0272-0590/89 $3.00
Copyright © 1989 by the Society of Toxicology.
All rights of reproduction in any form reserved.

gree to which the tumor changes mortality, and is robust with respect to survival differences among the groups. This test is developed using heuristic arguments for its justification, mathematical arguments for its derivation, and examples to indicate its ease of use. A study of the small sample properties of this test is given elsewhere (Bailer and Portier, 1988). In this study, it was observed that this modified test maintains the nominal false positive error rates when the treatment affects mortality.

METHODS

Data. The information obtained from most carcinogenicity experiments consists of the age at death of each animal and whether the tumor of interest was present or not at death. For the jth animal in the ith dose group, let t_{ij} denote the animals age at death and let $\delta_{ij} = 0$ if the animal did not have a tumor and let $\delta_{ij} = 1$ if one or more tumors was present, $i = 0, 1, \ldots, l$ and $j = 1, 2, \ldots, n_i$ (the analysis of multiple tumors in a single animal requires different procedures which are outside the focus of this paper). It is sometimes beneficial to have information on whether the death was a termination or a natural (nontermination) death. In the context of the present discussion, this information will not be used so notation for it is not defined.

The simplest framework for analyzing animal carcinogenicity experiments is (within each group) to consider the ratio of the number of animals with the tumor to the number of animals initially at risk. This ratio is usually referred to as the quantal estimate of response or simply quantal response. Let

$$x_i = \sum_{k=1}^{n_i} \delta_{ik}$$

denote the number of animals with the tumor in the group of animals given dose d_i of the test compound, $i = 0, 1, 2, \ldots, l$. The ratio x_i/n_i represents the quantal response.

A test of tumor incidence. The use of quantal response in the analysis of animal carcinogenicity experiments is based upon the assumption that all animals are at the same risk of getting the tumor over the duration of the experiment. Treatment can affect the survival patterns of the animals such that animals may die earlier in some treatment groups than in others. To use quantal response in this situation could result in an incorrect assessment of the carcinogenic potential of the test compound. As an example of an incorrect assessment of carcinogenicity, consider the following hypothetical carcinogen and an experiment with only two groups, a control group and a single-treated group.

Suppose the control tumor rate is such that the tumor does not appear before 90 weeks and appears in 10% of the animals very shortly after 90 weeks. Let us also assume that, unknown to the researcher, treatment increases this proportion to 30%. If all animals survived to 90 weeks and there were 50 animals in each group, it is expected that 5 out of 50 animals in the control group would have the tumor and 15/50 in the treated group would have the tumor shortly after 90 weeks, a comparatively strong finding of increased carcinogenesis. Now suppose treatment decreases survival such that only 60% of the treated animals survive to 90 weeks. Then only 9 of the 30 animals alive at this time would be expected to get the tumor. Without correcting for this survival difference, the quantal response methods would yield 9/50 as the proportion of animals with the tumor. This would represent a nonsignificant increase when compared against the control group. Thus, by not accounting for survival differences, it is possible to understate or reverse the carcinogenic effect. From this simple example it is easy to see why survival differences can play such an important role in the analysis.

Usually, several assumptions must be made to correct for survival differences in these experiments. Bailer and Portier (1988) have shown that the usual survival-adjusted carcinogenicity tests are sensitive to deviations from their respective assumptions. In addition, without modifying the design to incorporate additional terminations, it is not possible to test the validity of these assumptions. Thus, if existing methods are used, either the analyst must choose from a set of tests which are sensitive to untestable assumptions about the impact of tumor presence on survival or the experimenter must modify the experiment to include interim terminations at an increased cost.

There is yet one other approach to analyzing these data which has not been explored fully. This approach is a modification of simple quantal response methods that accounts for survival differences. Let R_i denote the risk of tumor development associated with an animal in dose group i that lives to the end of the study (at which point all live animals are intentionally killed). One way to approach the problem of treatment-related survival differences is to estimate what proportion of R_i is applicable to an animal which dies prior to the end of the study. In other words, modify the denominator, n_i, in the quantal estimate of response to more closely approximate the total number of animal years at risk.

This approach to the problem was suggested by Gart *et al.* (1979). They considered animals that die prior to some chosen time as not being at risk of getting the tumor and animals that die after that time as being at full risk. In mathematical terminology, if we let w_{ij} be the propor-

tion of the risk R_i associated with the jth animal in the ith group, then Gart *et al.* (1979) chose $w_{ij} = 0$ if $t_{ij} < C$ and $w_{ij} = 1$ if $t_{ij} \geqslant C$ where C is the cutoff time for inclusion in the analysis. They proposed the use of $C = 1$ (1 year) unless the first tumor is observed before 1 year, in which case they would use the time at which that tumor was observed. The analysis is then based on testing for trends as a function of treatment level using the modified proportions x_i/m_i, where m_i is the number of animals surviving past the cutoff time C in the ith treatment group. Bailer and Portier (1988) considered the false positive error rate and sensitivity of this test and found it still was moderately sensitive to survival differences between the groups. In what follows, a generalized version of this approach is developed and recommendations are given for choosing alternative weights (w_{ij}) to those of Gart *et al.* (1979).

Suppose instead of disregarding some animals as would be done by setting $w_{ij} = 0$, all animals were included but with some providing only fractional information. The question then becomes how much fractional information of the group risk R_i should be assigned to animals that die prior to study termination.

Each animal in a carcinogenicity experiment can be placed into one of four classes for the purposes of this analysis. The first two classes include animals which live to study termination and either have the tumor or do not have the tumor. Regardless of tumor status, all of these animals have contributed one full lifetime (or experiment time) of tumor experience. It follows then that their contribution to the estimate of R_i is complete and they should receive a weight of $w_{ij} = 1$. The third class of animals consists of those animals which die prior to study termination with the tumor present. Assuming tumors are irreversible (as most of the standard analyses assume), the early death of this animal would not affect its eventual tumor status. Thus, this animal can be considered as contributing an entire lifetime (experiment time) to the analysis and thus would also receive a weight of $w_{ij} = 1$.

To motivate the weights for the last class, animals that die early and are tumor-free, consider another example. Assume an animal dies at 1 year in a 2-year study and is tumor-free. This animal was at reduced risk of getting the tumor compared to animals that survived to study termination and should not be given a weight of 1. On the other hand, this animal did not get the tumor in the 1 year that it did live which does provide information on early tumor incidence. If the incidence of tumors was proportional to the age of the animal, this animal could be given a weight of $\frac{1}{2}$ implying it was at one-half as great a risk of getting the tumor as were animals that lived to term. When tumors seem to occur at a greater rate in older animals than in younger ones, this animal may be at less than one-half the risk of animals that survived the full 2 years. When attempts have been made to quantify the cumulative rate of tumor onset as a function of age, it generally has been found that tumors occur as a third to fourth order function of time (e.g., Doll, 1971; Portier *et al.*, 1986). This would suggest that the animal which dies without the tumor at 1 year in a 2-year study is at $(\frac{1}{2})^3$ or $(\frac{1}{2})^4$ the risk of animals which survived the full 2 years. Weights of this type represent one method for including all animals in the analysis. Other weighting schemes are possible.

Once weights have been chosen, the general procedure is as follows. For dose group i, the adjusted quantal estimate of response is given by

$$R_i^0 = \frac{x_i^0}{m_i^0},$$

where

$$x_i^0 = -\sum_{\substack{\text{All}\\ \text{Tumor-Bearing}\\ \text{Animals}}} w_{ij} = \sum_{j=1}^{n_i} \delta_{ij} w_{ij}$$

and

$$m_i^0 = \sum_{\text{All Animals}} w_{ij} = \sum_{j=1}^{n_i} w_{ij}.$$

The superscript 0 is a notational convenience to allow us to differentiate between the usual quantal response estimates and the modified quantal response estimates. These modified quantal estimates of response could then be used to replace the usual quantal response estimates, x_i/n_i. For example, a modified Cochran–Armitage test statistic (Cochran, 1954; Armitage, 1955) could be computed using the formula

$$\chi^2 = \left[\frac{M^0}{(M^0 - x_.^0)x_.^0}\right] \frac{\{\sum_{i=0}^{I} (x_i^0 - \gamma_i^0 x_.^0) d_i\}^2}{\sum_{i=0}^{I} \gamma_i^0 d_i^2 - \{\sum_{i=0}^{I} \gamma_i^0 d_i\}^2},$$

where

$$M^0 = \sum_{i=0}^{I} m_i^0, \quad x_.^0 = \sum_{i=0}^{I} x_i^0, \quad \gamma_i^0 = \frac{m_i^0}{M^0}.$$

The hypothesis of no increasing trend would be rejected at significance level α if χ^2 is larger than the upper $100\alpha\%$ of the χ^2 distribution with 1 degree of freedom. Bailer and Portier (1988) have shown that when the cumulative tumor incidence rate is a factorable function of age to the kth power, choosing the weights $w_{ij} = (t_{ij}/T)^k$ for tumor-free animals, where T is the length of the experiment, and $w_{ij} = 1$ for tumor-bearing animals results in an approximate test of tumor incidence which is robust to survival differences (nonfactorable hazards were not considered in this research). Their simulations and analytical results suggest the modified Cochran–Armitage trend test using $w_{ij} = (t_{ij}/T)^3$ for tumor-free animals and $w_{ij} = 1$ for tumor-bearing animals should function well in most exper-

imental situations even when $k \neq 3$. In what follows, an example is given where an analysis based on this procedure is compared to the Cochran–Armitage linear trend test (Cochran, 1954; Armitage, 1955), the modified trend test suggested by Gart *et al.* (1979), the Hoel–Walburg test (Hoel and Walburg, 1972) with intervals suggested by the National Toxicology Program, logistic regression (Dinse and Lagakos, 1983), and the life table test of Tarone (1975). These methods are frequently used to evaluate animal carcinogenicity experiments and are routinely used by the National Toxicology Program.

RESULTS

In a recent study of vinylcyclohexene diepoxide conducted by the National Toxicology Program (1989), reductions in survival were observed in treated female $B6C3F_1$ mice. The portion of the experiment in female mice utilized three treated groups of 50 animals dermally exposed to doses of vinylcyclohexene diepoxide of 25, 50, and 100 mg/ml. There was also a concurrent control group of 50 animals. Table 1 presents the ages at death (t_{ij}) and the presence or absence of alveolar/bronchiolar adenomas or carcinomas (δ_{ij}) in these 200 animals. It is clear from Table 1 that there is a large drop in survival in the highest dose group. The few remaining animals alive in this group at Week 85 were intentionally killed for humane reasons while the other groups continued until the end of the study at Week 105 ($T = 105$).

The unadjusted quantal response is 8% in control, 18% in the low-dose group, 22% in the mid-dose groups, and 14% in the high-dose group suggesting a downturn in response at the highest dose. The response at the highest dose is not significantly different from the control response using Fisher's exact test ($p = 0.262$). Using the weights given in Table 1 (e.g., $w_{ij} = (t_{ij}/T)^3$ or $=1$), the adjusted quantal response estimates are 9.6% (4/41.84) in control, 22% (9/40.97) in the low-dose group, 28.6% (11/38.49) in the mid-dose group, and 30.7% (7/22.81) in the high-dose group.

The modified Armitage linear trend test statistic can be calculated directly from Table 1. The results of this test along with the other tests mentioned above are presented in Table 2. The life table test attributes the most significance to the rejection of the hypothesis of no increased risk. Combined alveolar/bronchiolar adenomas or carcinomas have very little effect on survival (Portier *et al.*, 1986) indicating they are incidental. The extremely small p-value for the life table test is due to the large false positive error rate (Bailer and Portier, 1988) for this test when tumors are incidental.

The linear trend test and the Gart *et al.* (1979) modification to this test both result in a large p-value suggesting there is no increased carcinogenicity. With survival differences in the four groups of this magnitude, both of these tests exhibit a reduced false positive error rate relative to the nominal error rate (Dinse, 1985; Bailer and Portier, 1988) which results in a reduced sensitivity.

The Hoel–Walburg test has a larger p-value than does the logistic regression score test and the modified trend test using $k = 3$. As noted by Dinse (1985), this is due to the extreme early mortality in the highest dose group which results in no animals surviving until the terminal termination (the final interval in the Hoel–Walburg analysis). The logistic regression score test and the modified trend test gave equivalent p-values for these data. Since these tumors are probably incidental (Portier *et al.*, 1986), the score test is appropriate and should yield results essentially identical to those observed for the modified trend test.

DISCUSSION

Since the paper by Hoel and Walburg (1972), the analysis of the results of animal carcinogenicity experiments has steadily become more difficult mathematically. The methods that have been developed and advo-

TABLE 1

SURVIVAL DATA (t_{ij}), RISK WEIGHTS (w_{ij}), AND TUMOR INDICATOR (δ_{ij}) FOR COMBINED ALVEOLAR/BRONCHIOLAR ADENOMAS OR CARCINOMAS IN FEMALE B6C3F$_1$ MICE EXPOSED TO VINYLCYCLOHEXENE DIEPOXIDE

	Control			Low dose			Middle dose			High dose		
j	δ_{0j}	t_{0j}	w_{0j}	δ_{1j}	t_{1j}	w_{1j}	δ_{2j}	t_{2j}	w_{2j}	δ_{3j}	t_{3j}	w_{3j}
1		10	0.00		5	0.00		1	0.00		1	0.00
2		56	0.15		28	0.02		1	0.00		24	0.01
3		63	0.22		31	0.03		7	0.00		27	0.02
4		65	0.24		60	0.19		58	0.17		54	0.14
5		65	0.24		62	0.21		63	0.22		56	0.15
6		72	0.32		66	0.25		65	0.24		59	0.18
7		77	0.39		69	0.28		67	0.26		62	0.21
8		82	0.48		74	0.35	1	71	1.00	1	64	1.00
9		85	0.53		79	0.43	1	71	1.00		65	0.24
10		85	0.53		80	0.44		81	0.46		67	0.26
11		95	0.74		83	0.49		82	0.48		68	0.27
12		97	0.79		92	0.67		82	0.48		68	0.27
13		97	0.79		101	0.89		83	0.49		70	0.30
14		100	0.86		102	0.92	1	83	1.00	1	71	1.00
15		100	0.86		102	0.92		84	0.51		72	0.32
16		101	0.89		102	0.92		85	0.53		72	0.32
17		102	0.92	1	102	1.00		86	0.55		72	0.32
18		103	0.94	1	102	1.00	1	87	1.00		75	0.36
19	1	103	1.00		104	0.97		93	0.70		75	0.36
20		103	0.94	1	105	1.00		96	0.76		75	0.36
21	1	105	1.00	1	105	1.00	1	96	1.00		76	0.38
22	1	105	1.00	1	105	1.00		97	0.79		77	0.39
23	1	105	1.00	1	105	1.00		97	0.79		77	0.39
24		105	1.00	1	105	1.00		98	0.81		78	0.41
25		105	1.00	1	105	1.00		98	0.81		78	0.41
26		105	1.00	1	105	1.00		99	0.84		78	0.41
27		105	1.00		105	1.00	1	100	1.00		79	0.43
28		105	1.00		105	1.00		101	0.89		79	0.43
29		105	1.00		105	1.00		101	0.89		79	0.43
30		105	1.00		105	1.00		103	0.94		79	0.43
31		105	1.00		105	1.00		104	0.97		80	0.44
32		105	1.00		105	1.00		104	0.97		81	0.46
33		105	1.00		105	1.00		104	0.97		83	0.49
34		105	1.00		105	1.00		104	0.97		83	0.49
35		105	1.00		105	1.00	1	104	1.00		83	0.49
36		105	1.00		105	1.00	1	105	1.00	1	83	1.00
37		105	1.00		105	1.00	1	105	1.00		83	0.49
38		105	1.00		105	1.00	1	105	1.00		83	0.49
39		105	1.00		105	1.00	1	105	1.00		85	0.53
40		105	1.00		105	1.00		105	1.00		85	0.53
41		105	1.00		105	1.00		105	1.00	1	85	1.00
42		105	1.00		105	1.00		105	1.00	1	85	1.00
43		105	1.00		105	1.00		105	1.00		85	0.53
44		105	1.00		105	1.00		105	1.00	1	85	1.00
45		105	1.00		105	1.00		105	1.00	1	85	1.00
46		105	1.00		105	1.00		105	1.00		85	0.53
47		105	1.00		105	1.00		105	1.00		85	0.53
48		105	1.00		105	1.00		105	1.00		85	0.53
49		105	1.00		105	1.00		105	1.00		85	0.53
50		105	1.00		105	1.00		105	1.00		85	0.53
	4		41.84	9		40.97	11		38.49	7		22.81

Note. No entry for δ_{ij} implies $\delta_{ij} = 0$.

TABLE 2

TREND TESTS OF TUMOR INCIDENCE FOR ALVEOLAR/BRONCHIOLAR ADENOMAS OR CARCINOMAS IN FEMALE $B6C3F_1$ MICE EXPOSED TO VINYLCYCLOHEXENE DIEPOXIDE

Test	p-Value
Life table test	$p < 0.001$
Cochran–Armitage linear trend test	$p = 0.528$
Gart *et al.* linear trend test	$p = 0.412$
Hoel–Walburg (incidental tumor) test	$p = 0.075$
Logistic regression score test	$p = 0.033$
Modified trend test ($k = 3$)	$p = 0.034$

cated by statisticians have progressed to the stage where the researcher has very little understanding of what is being done and must rely upon the statistician and the computer to determine if the results differ significantly from random fluctuation in tumor rates. The modified trend test proposed in this manuscript is a test which the researcher can do on a sheet of paper. Under simple assumptions, the test is a valid test of lifetime (or experiment time) tumor incidence which is seemingly unaffected by group differences in survival (Bailer and Portier, 1988).

As discussed in the introduction, all methods for analyzing data from carcinogenicity experiments rely upon untestable assumptions. The one critical, untestable assumption in the use of the survival-adjusted quantal-response method presented in the text is that the cumulative tumor incidence function must be a factorable function of dose and time to some power, k. In a series of Monte Carlo simulation experiments, Bailer and Portier (1988) have shown that by using $k = 3$, the test remains valid (i.e., possesses nominal false positive error rate) even when the true value of k is as small as 1 or as large as 5. Alternative choices for k may be desired for some tumors. One source of alternative values is the analysis of historical control animals of Portier *et al.* (1986). The estimates for the parameter β presented in Tables 3 through 6 of that manuscript are directly comparable to the weighting exponent k. For combined alveolar/bronchiolar adenomas or carcinomas in Female $B6C3F_1$ mice, this parameter was estimated as 3.05.

The modified trend test is still a proposal. Before the usual tests for carcinogenicity can be replaced by this test, additional research is needed. Although many of the usual tests of carcinogenicity use factorable hazards, this test requires a specific form with age^k as one of the factors. It is unknown what effect other models with different factorable hazards and unfactorable hazards would have on the false-positive error rate and sensitivity of this test. This issue should be addressed before the test is used routinely.

ACKNOWLEDGMENTS

We thank Dr. Kenneth Brown and Dr. Joseph Hasemen for their input into the use of the modified trend test and Dr. Gregg Dinse for his careful reading of this manuscript. Also, we thank Dr. Haseman for the data from the carcinogenicity experiment on vinylcyclohexene diepoxide.

REFERENCES

ARMITAGE, P. (1955). Tests for linear trends in proportions and frequencies. *Biometrics* **11**, 375–386.

ARMITAGE, P., AND DOLL, R. (1954). The age distribution of cancer and a multistage theory of carcinogenesis. *Brit. J. Cancer* **8**, 1–12.

BAILER, A. J., AND PORTIER, C. J. (1988). Effects of treatment-induced mortality and tumor-induced mortality on tests for carcinogenicity in small samples. *Biometrics* **44**, 417–432.

COCHRAN, W. G. (1954). Some methods for strengthening the common χ^2 tests. *Biometrics* **10**, 417–451.

DEWANJI, A., AND KALBFLEISCH, J. D. (1986). Nonparametric methods for survival/sacrifice experiments. *Biometrics* **42**, 325–341.

DINSE, G. E. (1985). Testing for a trend in tumor prevalence rates: I. Nonlethal tumors. *Biometrics* **41**, 751–770.

DINSE, G. E., AND LAGAKOS, S. W. (1983). Regression analysis of tumor prevalence data. *Appl. Stat.* **32**, 236–248.

DOLL, R. (1971). The age distribution of cancer: Implications for models of carcinogenesis. *J. R. Stat. Soc. Ser. A* **37,** 133–166.

GART, J. J., CHU, K., AND TARONE, R. E. (1979). Statistical issues in the interpretation of chronic bioassay tests for carcinogenicity. *J. Natl. Cancer Inst.* **62,** 957–974.

HOEL, D. G., AND WALBURG, H. E. (1972). Statistical analysis of survival experiments. *J. Natl. Cancer Inst.* **49,** 361–372.

MCKNIGHT, B. (1988). A guide to the statistical analysis of long-term carcinogenicity assays. *Fundam. Appl. Toxicol.* **10,** 355–364.

MCKNIGHT, B., AND CROWLEY, J. J. (1984). Tests for differences in tumor incidence based on animal carcinogenesis experiments. *J. Amer. Statistical Assoc.* **79,** 639–648.

National Toxicology Program (1989). *NTP Technical Report No. 362 on the Toxicology and Carcinogenesis Studies of 4-Vinyl-1-cyclohexene Diepoxide in F344/N Rats and B6C3F$_1$ Mice.* National Toxicology Program.

PORTIER, C. J., AND DINSE, G. E. (1987). Semiparametric analysis of tumor incidence rates in survival/sacrifice experiments. *Biometrics* **43,** 107–114.

PORTIER, C. J., HEDGES, J. C., AND HOEL, D. G. (1986). Age-specific models of mortality and tumor onset for historical control animals in the National Toxicology Program's carcinogenicity experiments. *Cancer Res.* **46,** 4372–4378.

TARONE, R. E. (1975). Tests for trend in life table analysis. *Biometrika* **62,** 679–682.