FUNDAMENTAL AND APPLIED TOXICOLOGY 6, 44–52 (1986)

Logistic Regression Analysis of Incidental-Tumor Data from Animal Carcinogenicity Experiments

GREGG E. DINSE AND JOSEPH K. HASEMAN

Biometry and Risk Assessment Program, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina 27709

Logistic Regression Analysis of Incidental-Tumor Data from Animal Carcinogenicity Experiments. DINSE, G. E., AND HASEMAN, J. K. (1986). *Fundam. Appl. Toxicol.* **6**, 44–52. Survival differences can have a substantial impact on the statistical comparison of tumor development in control and treated animals and thus should be taken into account routinely in the analysis of carcinogenicity data from laboratory experiments. However, the appropriate survival adjustment depends on whether the tumor of interest is fatal or incidental. The usual analysis of incidental tumors, which adjusts for survival by stratifying the animals according to age at death, has various shortcomings. Alternatively, logistic regression methods allow a continuous survival adjustment and furnish a convenient framework for solving many of the problems associated with the age-stratified approach of grouping the data into time intervals. Logistic regression substitutes modeling the prevalence function for the arbitrary choice of time intervals, providing a survival adjustment (when the model holds) even when differential mortality might increase the bias or decrease the sensitivity of interval-based methods. The logistic analysis also can incorporate covariables which, if ignored, might confound the interpretation of the data. Several examples illustrate these potential advantages of basing the analysis of incidental tumors on logistic regression techniques. © 1986 Society of Toxicology.

The animal carcinogenesis experiment is an important instrument for assessing environmental risk. These studies typically involve both sexes of two rodent species, usually mice and rats, and compare a control group of unexposed animals with one or more treated groups that are exposed to increasing dose levels of a test chemical. There are commonly 50 animals in each treated and control group, for each of the four sex/species combinations. We will focus on long-term survival/sacrifice studies that observe animals for the majority of their normal lifespan. Often the study is terminated at the end of some fixed period of time (e.g., 2 years), at which point all remaining animals are sacrificed. Pathologists examine each animal at death, both grossly and microscopically, and record the age at death (survival time) and a list of the various tumors and other lesions discovered at necropsy.

One of the main objectives of a carcinogenicity experiment is to compare control and dosed groups of animals with respect to tumor development. As a general rule, the statistical analysis should take survival information into account because differential mortality (across groups) can have a substantial impact on the interpretation of the data (see, for example, Hoel and Walburg, 1972; Gart *et al.*, 1979; Peto *et al.*, 1980). As the chances of developing (and subsequently dying as a result of) a tumor can vary with age, an analysis that does not adjust for age (survival) can produce misleading conclusions when the control and dosed groups have unequal mortality patterns.

For example, suppose that treatment with a certain chemical has no effect on tumor development, but that the chemical is toxic enough that the dosed animals die much earlier than the controls. An analysis that does

0272-0590/86 $3.00

not adjust for survival might falsely declare that the chemical is protective against cancer simply because the dosed animals are dying before tumors have time to develop. Conversely, if a chemical is both toxic and carcinogenic, then early deaths might preclude enough tumors to yield similar lifetime tumor incidences in the control and dosed groups. Therefore, a true carcinogenic response can be masked if survival differences are not taken into account. In fact, even when the groups have identical survival experiences, an analysis that ignores information on the age at death can be less efficient than a survival-adjusted analysis, even though the former is still valid in this case (Ryan, 1985). See Table 3 of Gart *et al.* (1979) for a list of interpretations of an unadjusted analysis in light of survival information in these (and other) situations.

The appropriate method of adjusting for survival depends on a tumor's role in causing death, i.e., its *context of observation* (Peto *et al.*, 1980). A tumor discovered at necropsy is considered *fatal* if it was directly or indirectly responsible for the animal's death and *incidental* if it was revealed merely as the result of a death from an unrelated cause. The distinction between fatal and incidental tumors is important because it is essential to distinguish between a chemical that reduces survival by shortening the time to tumor onset or the time to death following tumor onset (a real carcinogenic effect), and one that also reduces survival, but for which tumors are observed earlier simply because animals are dying of competing causes (a noncarcinogenic effect). One of the most comprehensive discussions of the analysis of animal carcinogenicity data is given by Peto *et al.* (1980), who emphasize the importance of adjusting for survival and the need to determine the context of observation for each tumor discovered at necropsy. Kodell *et al.* (1982) also advocate the collection of individual cause-of-death data and illustrate its usefulness.

If the tumor type is so lethal (i.e., progresses so rapidly) that all occurrences are fatal, the control and dosed groups should be compared via *life-table* methods, which focus on the age-specific hazard rates of death due to the tumor (i.e., the proportions of animals dying from the tumor at a given time, among those alive just prior to that time). Conversely, if the tumor type is strictly nonlethal, so that tumor development has no effect on the risk of death, the groups should be compared via *prevalence* methods, which focus on the age-specific tumor prevalence rates (i.e., the proportions of live animals having the tumor). See Hoel and Walburg (1972) for a discussion of the standard life-table and prevalence analyses. If the tumor type is of moderate lethality, so that some occurrences are fatal and others are incidental, the analysis can be based on a combination of life-table and prevalence methods, as long as each tumor's context of observation is identified (Peto, 1974; Peto *et al.*, 1980). In this case, a life-table analysis is applied to all of the data and a prevalence analysis is applied to only the subset of animals dying from causes other than the tumor. These two analyses can be interpreted separately or they can be combined via Mantel–Haenszel methods (Mantel and Haenszel, 1959) to give an overall test for a dose effect. A combined analysis makes sense if the test chemical elevates both the tumor prevalence rate and the hazard rate of death due to the tumor, but might have little sensitivity (power) if one rate increases and the other rate decreases (Lagakos, 1982). For additional discussions of carcinogenicity testing in animal experiments see, for example, Turnbull and Mitchell (1978), Mitchell and Turnbull (1979), Haseman (1984), Lagakos and Louis (1985), and McKnight and Crowley (1984).

In this paper, we concentrate on the analysis of incidental tumors. The associated prevalence analysis involves all of the animals if the tumor type is strictly nonlethal or just the subset of animals dying from other causes if some occurrences of the tumor are fatal. We compare the usual Mantel–Haenszel method of adjusting for survival (i.e., by stratifying on age at death) with an alternative survival-adjusted analysis based on logistic regression

techniques. Our purpose is to emphasize and illustrate (via example) some of the potential advantages of the logistic regression analysis relative to the standard analysis.

METHODS

The statistical techniques traditionally used by the National Cancer Institute in their Carcinogenesis Bioassay Testing Program were Fisher's exact test for pairwise comparisons of each dose group with the control group and the Cochran–Armitage test (Cochran, 1954; Armitage, 1955) for detecting a dose-response trend (see Gart *et al.*, 1979). These procedures compare the lifetime tumor incidence rates (i.e., the overall proportions of tumor-bearing animals) in the control and dosed groups. Whereas these methods do not require the context of observation for each animal, they also make no adjustments for possible dose-related survival differences which could bias the results. One simple modification that makes a crude adjustment for differential mortality is to ignore the animals dying before the first death with a tumor. Though in some cases the comparison of lifetime tumor rates among animals surviving until the first death of a tumor-bearing animal might be satisfactory, generally a more sophisticated survival adjustment is necessary to guard against bias.

The National Toxicology Program (NTP) of the U.S. Department of Health and Human Services currently is responsible for the Carcinogenesis Bioassay Testing Program and routinely incorporates formal survival-adjusted methods in their statistical analyses of carcinogenicity data. The NTP's incidental-tumor test adjusts for survival by stratifying on age at death, as described by Hoel and Walburg (1972) and Peto *et al.* (1980). Henceforth we refer to this standard survival-adjusted prevalence test as the Hoel–Walburg test.

The first step in computing the Hoel–Walburg test is to select a set of intervals with which to partition the survival times. If any sacrifices are performed, such as at the end of the study or at specified interim times, all animals killed in each such sacrifice period are placed in their own separate interval. Second, specific to each dose group, the *observed* number of animals dying with a tumor in a particular time interval is compared with the number of tumor-bearing animals *expected* to die in that interval under the hypothesis that dose has no effect on tumor prevalence. Third, the differences between the observed and expected numbers of deaths with a tumor are combined across intervals via the methods of Mantel and Haenszel (1959) to yield an overall survival-adjusted test for a dose effect on tumor prevalence rates. The approximate statistical significance of an apparent dose effect can be obtained by comparing the Hoel–Walburg statistic to published tables of the percentage points for the standard normal (Gaussian) distribution. The Hoel–Walburg test can be employed for assessing dose–response trends or for making pairwise comparisons of each dose group with the control group. See Haseman (1984) for computational details and illustrative sample calculations.

Dinse and Lagakos (1983) propose an alternative prevalence trend test for incidental tumors that is based on logistic regression techniques (Cox, 1970). They use the logistic distribution to model (express) tumor prevalence as an explicit function of dose and age, as well as possible confounding covariables such as gender, weight, and cage location. The logistic analysis specifies a regression coefficient for dose that, within the parametric constraints of the model, measures the effect of dose on tumor prevalence. Dinse and Lagakos (1983) focus on the so-called likelihood score test as a means of detecting a trend in tumor prevalence rates. For a detailed discussion of the logistic regression analysis, see Dinse and Lagakos (1983), and for descriptions of computer programs for performing the analysis, see Amini and Palka (1985) and Dinse (1986).

The Hoel–Walburg analysis can be viewed as a special case of a general logistic analysis; that is, the likelihood score test under a certain logistic model essentially reduces to the Hoel–Walburg test (Dinse and Lagakos, 1983; Day and Byar, 1979). Mainly, the tests differ in how they adjust for age. The Hoel–Walburg test implicitly assumes that tumor prevalence is a step function of age (i.e., constant over time intervals), which implies that tumors can occur only at the endpoints of certain time intervals. In contrast, the logistic score test assumes that tumor prevalence is a smooth function of age, which implies that tumors can occur at any time during the study. Therefore, one advantage of the logistic score test is the fact that a smooth prevalence function is more biologically reasonable than a step function. Furthermore, the logistic regression analysis adjusts for survival without having to form time intervals or group the data, and it easily incorporates available information on covariables which might confound the usual prevalence analysis.

RESULTS

An incidental-tumor test based on a regression model, such as the logistic score test proposed by Dinse and Lagakos (1983), enjoys several advantages over the standard Mantel–Haenszel type tests proposed by Hoel and Walburg (1972) and Peto *et al.* (1980). In this section, we discuss and illustrate three important advantages of the logistic regression approach.

The logistic score test adjusts for survival by directly incorporating the individual death times, which circumvents some of the problems associated with the standard interval-based tests. The Hoel–Walburg test adjusts for

survival by stratifying on age at death, and thus time intervals must be selected for grouping the data. A reasonable interval must be wide enough to include a sufficient number of animals for a meaningful treatment comparison, and yet narrow enough to minimize the bias due to violations of the assumption that the tumor prevalence is constant within each interval. In practice, a small set of fixed-length intervals is often chosen, though the exact number and length of the intervals are somewhat arbitrary. Alternatively, Peto *et al.* (1980) propose a formal method of interval selection that produces intervals of varied number and length, depending on the observed data. Unfortunately, not only is there no widely accepted "best" method of selecting time intervals, but different sets of intervals can lead to different conclusions. Hence, one potential advantage of the logistic analysis is that it avoids the arbitrariness of choosing a set of time intervals, as well as the problem of reconciling the discrepant results that might arise from different choices of intervals.

For illustration purposes, consider the data originally reported by Hoel and Walburg (1972) on the incidence of lung tumors in 144 male RFM mice. There are two groups and no covariables in this data set. The ages at death range from 45 to 1008 days. We calculated the Hoel–Walburg statistic for eight sets of fixed-length intervals. The interval lengths and the associated two-sided significance levels are: 1 month ($p = 0.261$), 2 months ($p = 0.063$), 3 months ($p = 0.043$), 4 months ($p = 0.029$), 6 months ($p = 0.015$), 1 year ($p = 0.016$), 2 years ($p = 0.011$), and 3 years ($p < 0.001$). Note that a single 3-year interval contains all of the death times. Thus the corresponding Hoel–Walburg test makes no survival adjustment and can be viewed as an approximation to Fisher's exact test, which also yields $p < 0.001$. Hoel and Walburg (1972) used 100-day intervals, which give a significance level of $p = 0.024$, whereas the interval selection method of Peto *et al.* (1980) yields a set of 11 variable-length intervals and a significance level of $p = 0.126$. Clearly the statistical significance of the observed data depends on the choice of time intervals.

Alternatively, we calculated the logistic score test under models that specified tumor prevalence as a logistic function of dose and either a linear, quadratic, or cubic function of age. The significance levels corresponding to these three models are: linear ($p = 0.025$), quadratic ($p = 0.021$), and cubic ($p = 0.022$). The results of the logistic analyses are consistent, regardless of the smooth (polynomial) function used to model age, and thus the arbitrariness associated with choosing time intervals does not necessarily carry over to the choice of the smooth function of age. If the tumor type is irreversible and nonlethal, the prevalence rates should increase with age and typically a linear model in age will suffice (Dinse, 1985). Otherwise, age-squared or age-cubed terms can be included initially and removed if the regression analysis suggests that their contribution is negligible (i.e., not statistically significant).

Another difficulty encountered with an interval-based test is the problem of handling differential mortality, which might arise from treatment toxicity or unbalanced sacrificing. Animals that die in a time interval that contains deaths from only a single treatment group make no contribution at all to the Hoel–Walburg statistic. Therefore, when the groups have different mortality patterns, the choice of time intervals is complicated further. In extreme cases, either some of the intervals must be made wider than usual in order to include animals from more than one group, or else some of the animals must be ignored. The first option runs the risk of introducing a bias and the second option reduces the effective sample size. Hence, another potential advantage of the logistic analysis is that it can use all of the data, even when the groups have different survival experiences, and thus generally will be either less biased, or else more powerful, than the Hoel–Walburg test.

As an example of how the logistic score test might identify an effect that the Hoel–Walburg test could miss, consider the data reported in

Table 1. Table 1 summarizes the survival experience and the incidence of tumors of the subcutaneous tissue for the 100 male F344 rats in the control and high-dose groups of the NTP carcinogenesis bioassay of 1,2-dibromoethane (NTP, 1982). Of the 50 rats in the control group, 38 (76%) survived until the end of the 2-year experiment and were sacrificed between Weeks 104 and 106. Conversely, none of the 50 rats in the high-dose group survived to the terminal sacrifice; in fact they all died by Week 89, including five that were killed at that time because of their moribund condition. These five moribund animals were killed for humane reasons and, because they would have died in the very near future, their deaths were treated as natural rather than sacrificial for the purposes of analysis. There was little overlapping survival between the groups, which makes it difficult to choose intervals containing death times from both groups.

TABLE 1

SURVIVAL AND INCIDENCE OF FIBROMAS OF THE SUBCUTANEOUS TISSUE FOR 100 MALE F344 RATS IN THE NTP STUDY OF 1,2-DIBROMOETHANE

	Control group		High-dose group	
Age at death (in weeks)	No tumor	Tumor	No tumor	Tumor
43	0	0	1	0
50	0	0	1	0
53	0	0	1	0
56	0	0	1	0
62	0	0	1	0
63	0	0	1	0
64	0	0	2	0
67	0	0	2	0
68	0	0	2	0
69	0	0	1	0
70	0	0	4	0
71	0	0	1	0
74	0	0	1	0
76	0	0	2	0
77	1	0	0	1
78	0	0	1	0
80	0	0	4	1
81	0	0	2	0
82	0	0	2	1
83	0	0	1	0
84	0	0	2	1
85	0	0	1	1
86	0	0	1	1
87	0	0	1	1
88	1	0	1	0
89	1	0	5	1
90	1	0	0	0
93	2	0	0	0
97	1	0	0	0
99	1	0	0	0
102	1	0	0	0
103	3	0	0	0
104[a]	16	3[b]	0	0
106[a]	18	1	0	0
Total	46	4	42	8

[a] All animals with death times of 104 or 106 weeks were sacrificed.

[b] One of these animals had a fibrosarcoma rather than a fibroma.

An unadjusted analysis compares the lifetime tumor incidence rates in the control (8%) and high-dose (16%) groups; Fisher's exact test yields a one-sided significance level of $p = 0.178$. If we ignore the 21 animals (all in the high-dose group) that die before the first tumor is observed at Week 77, as suggested by Gart *et al.* (1979), then Fisher's exact test yields $p = 0.024$.

In their 2-year studies, the NTP usually forms one interval for the animals sacrificed at the end of the study and four intervals (measured in weeks) for the animals dying of natural causes: 0–52, 53–78, 79–92, and 93–103. The first of these intervals is meant to contain the early deaths (i.e., those occurring in the first year) and each of the last three intervals is meant to contain roughly a third of the natural deaths occurring in the second year. In the current example, this choice of intervals essentially cuts the effective sample size in half, as there are only 52 rats that die in intervals containing deaths from both groups. The Hoel–Walburg test based on these intervals yields a significance level of $p = 0.156$. If we modify the usual NTP intervals so that the middle two intervals are 79–88 and 89–103, which compares the five moribund rats in the high-dose group with all control rats dying naturally from that point on, the analysis still ignores 40 rats and the Hoel–

Walburg test gives $p = 0.181$. Similarly, if all rats dying at or after Week 89 (including those sacrificed at the terminal kill) are included in the final interval, the analysis only ignores two rats, although the Hoel–Walburg statistic remains about the same ($p = 0.184$). The interval selection method of Peto *et al.* (1980) produces three intervals (0–76, 77–103, and 104–106), only the middle of which contains deaths from both groups. This analysis ignores 59 rats and the associated Hoel–Walburg test gives $p = 0.023$.

The logistic score test, based on a model with simply a linear age term, yields a significance level of $p = 0.003$. A model with a linear age term was deemed sufficient because the regression coefficient for the age-squared term in a model with a quadratic function of age was not statistically significant.

In the unadjusted analysis, there is no evidence that the lifetime tumor incidence rates are different in the two groups. Similarly, the Hoel–Walburg test based on the NTP intervals, or simple modifications of these intervals, provides no evidence that the age-specific prevalence rates differ. Fisher's exact test, applied to the subsample of animals dying at or after the first death of a tumor-bearing animal, and the Hoel–Walburg test, applied with the intervals obtained by the method of Peto *et al.* (1980), both yield marginal evidence of higher age-specific prevalence rates in the high-dose group. The NTP called 1,2-dibromoethane a carcinogen in male rats on the basis of tumors of the nasal cavity, the circulatory system and the reproductive system, but they did not detect an effect on tumor development in the subcutaneous tissue. The logistic score test, on the other hand, provides fairly strong evidence of an increased prevalence of subcutaneous-tissue tumors in the high-dose group, which supports the general conclusion of carcinogenicity made by the NTP.

Finally, the Hoel–Walburg test makes no provision for covariables which might confound the interpretation of the study. That is, differences in tumor prevalence rates that are attributed to the chemical being tested might in fact be explained entirely by differences in how important covariables (explanatory factors) are distributed across the control and dosed groups. Therefore, a third advantage of the logistic analysis is that it adjusts for covariables. We have created a set of hypothetical data, provided in Table 2, to illustrate such an adjustment. There are 25 animals in each of three treatment groups, where the dose levels are 0, 1, and 2. Each animal has an age at death, a tumor response indicator, and two covariables. Age at death is measured in weeks. The response indicator equals 1 if the tumor is present at death and 0 otherwise. The first covariable represents initial weight at the beginning of the study (e.g., at 6 weeks of age) and is measured in grams. The second covariable represents the row of the rack in which the animal's cage is housed and takes a value of 1, 2, or 3 to signify top, middle, or bottom, respectively. All animals with an age of 104 weeks were sacrificed at a 2-year terminal kill.

An analysis that does not adjust for survival or covariables suggests a definite (positive) trend in lifetime tumor incidence rates. The lifetime tumor rates are 8% (2/25) in the control group, 20% (5/25) in the low-dose group, and 36% (9/25) in the high-dose group, and the Cochran–Armitage test produces a one-sided significance level of $p = 0.008$. If we ignore the four animals that die before the first tumor is observed at Week 82, the Cochran–Armitage test gives a value of $p = 0.005$. Similarly, the Hoel–Walburg test based on the intervals selected by the method of Peto *et al.* (1980) yields $p = 0.008$, and the Hoel–Walburg test based on the NTP intervals yields $p = 0.009$. In fact, the logistic score test based on a model that specifies prevalence as a logistic function of dose and a linear term in age also yields $p = 0.008$.

As the mortality patterns are similar in all three groups, an adjustment for survival is not really necessary and thus the analyses that adjust for age (but not covariables) lead to the same conclusion as drawn from the unadjusted analysis: tumor prevalence increases as dose increases. However, both initial weight and cage location are related to tumor response and treatment group. That is, both the tumor-

TABLE 2

HYPOTHETICAL DATA ON DOSE, AGE AT DEATH, TUMOR RESPONSE, INITIAL WEIGHT, AND CAGE LOCATION (ROW) FOR 75 ANIMALS[a]

High-dose group				Low-dose group				Control group			
Age	Resp	Wt	Row	Age	Resp	Wt	Row	Age	Resp	Wt	Row
82	1	99	1	81	0	94	3	88	0	91	2
85	1	97	1	83	0	98	3	101	0	95	2
99	1	96	1	100	0	95	3	104	0	94	2
101	0	95	1	104	0	95	3	104	0	97	2
104	0	92	1	104	0	93	3	104	0	95	2
87	1	99	2	87	1	97	1	91	0	97	3
91	0	96	2	94	1	96	1	92	0	96	3
91	0	98	2	102	0	92	1	97	0	95	3
104	0	91	2	104	0	94	1	104	0	93	3
104	1	95	2	104	1	96	1	104	0	93	3
81	0	96	3	90	0	98	2	87	0	94	1
93	0	97	3	92	0	93	2	100	0	94	1
97	0	93	3	94	0	92	2	104	0	95	1
102	0	94	3	104	0	95	2	104	1	97	1
104	0	95	3	104	0	93	2	104	1	98	1
81	0	92	1	80	0	91	3	96	0	95	2
97	1	98	1	100	0	98	3	100	0	96	2
103	1	96	1	104	0	95	3	104	0	95	2
104	0	96	1	104	0	95	3	104	0	92	2
104	1	95	1	104	0	94	3	104	0	97	2
93	0	96	2	86	0	92	1	83	0	95	3
98	0	92	2	98	1	97	1	104	0	95	3
103	0	98	2	104	0	94	1	104	0	94	3
104	0	94	2	104	0	95	1	104	0	95	3
104	1	99	2	104	1	99	1	104	0	96	3

[a] Age at death is measured in weeks and all animals alive at 104 weeks were sacrificed at that time. Tumor response (Resp) equals 1 if the tumor is present and 0 otherwise. Initial weight (Wt) is measured in grams. Cage location (Row) equals 1 if housed on the top row, 2 if housed in the middle row, and 3 if housed on the bottom row.

bearing animals and the high-dose animals tend to have higher initial weights and tend to be housed in the first row of the rack more often. Therefore, the association between higher dose levels and increased prevalence rates can be explained, at least to some extent, by the association between initial weight, cage location, and tumor response. If the prevalence is modeled as a logistic function of initial weight and row, as well as dose and age, the apparent positive trend in prevalence rates with increasing dose disappears. The logistic score test based on a model with linear terms in dose, age, initial weight, and row yields a significance level of $p = 0.113$. Quadratic and cubic terms in age were not significant. In fact, the linear term in age was not even significant, though the logistic score test based on a model with coefficients for only dose, initial weight, and row gave virtually the same result ($p = 0.115$). Therefore, an analysis that does not account for potential confounders can produce conclusions that differ from an analysis that adjusts for information on these explanatory factors.

DISCUSSION

We have focused entirely on incidental tumors. In practice, pathologists consider certain tumor types to be nonlethal; hence all occurrences of these tumors are incidental by defi-

nition and individual contexts of observation are unnecessary. However, if the tumor type is moderately lethal, some occurrences might be fatal and others incidental. In this case, the incidental-tumor tests should be applied only to the subsample of animals dying from causes other than the tumor, which is not possible without accurate assessments of the context of observation for each tumor. Until recently, the NTP has not collected individual data on cause of death. Rather, the NTP routinely reports the results of three analyses for each tumor type: one that does not adjust for survival, one that treats all tumors found at natural death as fatal, and one that treats all tumors as incidental. Frequently all three analyses lead to the same conclusions, but occasionally the results differ and the decision becomes more complicated. As part of its new modified pathology protocol, however, the NTP requests that pathologists attempt to determine the context of observation for each tumor observed in an NTP study. While many pathologists are skeptical about the reliability of forced cause-of-death assessments, Peto *et al.* (1980, p. 330) report that of over 4500 tumors discovered in a large experiment, 94% were classified as either definitely fatal or definitely incidental, despite the initial reservations of the pathologists involved.

One of the advantages of the logistic analysis is that, by virtue of being a parametric approach, it makes a continuous survival adjustment rather than having to stratify on age by grouping the data into time intervals. This advantage will probably be the greatest when mortality patterns differ across groups and yet there is still some overlapping survival. In this situation, the logistic analysis will make more efficient use of the available data than the interval-based analysis. If the mortality patterns are similar in all groups, and there are no important covariables, both analyses should perform equally as well (Dinse, 1985). At the opposite extreme, if there is no overlap in survival, the Hoel–Walburg test is indeterminant and though the logistic score test might be calculable, the survival adjustment would rely heavily on untestable parametric assumptions. Note, however, that as long as there is some overlapping survival, the regression framework allows the logistic analysis to check certain of these assumptions.

The data in Table 1 are fairly extreme, in the sense that only three control rats die at or before the last death time in the high-dose group. Nevertheless, there is still evidence that the effect detected by the logistic score test is real. Suppose we group the control rats sacrificed at 104 and 106 weeks and treat this pooled number of tumor-bearing rats as a binomial observation at Week 105. Under the binomial model, the estimated prevalence at week 105 is 10.5% (4/38) and the exact 99% upper confidence bound on that estimate is 27.6%. If we then separately model the prevalence in the high-dose group by a logistic function with just a linear term in age, the estimated prevalence at Week 89 is 37.4% and the corresponding approximate 99% lower confidence bound is 28.1%. As the confidence bounds do not overlap, we can conclude that the prevalence in the high-dose group at Week 89 is significantly greater than the prevalence in the control group at Week 105. If tumors of the subcutaneous tissue are irreversible and do not alter the risk of death, their prevalence cannot decrease with age, and hence the prevalence in the control group at Week 89 cannot exceed that at Week 105. Thus, there is evidence that at Week 89 the prevalence in the high-dose group significantly exceeds the prevalence in the control group and, had the high dose not been so toxic, this difference also would have been evident at the end of the study.

In conclusion, statistical analyses of animal carcinogenicity data should routinely adjust for survival because differential mortality across treatment groups can bias lifetime (unadjusted) dose–response comparisons. However, the proper method of taking survival into account depends on whether tumors are fatal or incidental. When analyzing incidental tumors, the logistic approach permits an adjustment for survival without requiring the data to be grouped into time intervals, as required by the age-stratified Mantel–Haenszel

tests proposed by Hoel and Walburg (1972) and Peto *et al.* (1980). As a result, the logistic score test avoids the arbitrariness associated with choosing the time intervals, as well as the problem of explaining the different conclusions obtained from different sets of intervals. Moreover, the logistic score test frequently can take survival into account even when differential mortality makes it difficult for the Hoel–Walburg test to adjust for survival without introducing bias or insensitivity. Finally, the logistic analysis provides a framework for incorporating information on important covariables which could confound analyses that ignore these explanatory factors.

REFERENCES

AMINI, S. B., AND PALKA, D. L. (1985). Application of the SAS system to regression analysis of tumor prevalence data. *In* "Proceedings of the Tenth Annual SAS Users Group International Conference," Las Vegas, Nev.

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

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

COX, D. R. (1970). *The Analysis of Binary Data.* Methuen, London.

DAY, N. E., AND BYAR, D. P. (1979). Testing hypotheses in case-control studies—equivalence of Mantel–Haenszel statistics and logit score tests. *Biometrics* **35,** 623–630.

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

DINSE, G. E. (1986). LOPRAN: A program for logistic prevalence analysis. Unpublished manuscript.

DINSE, G. E., AND LAGAKOS, S. W. (1983). Regression analysis of tumor prevalence data. *Appl. Stat.* **32,** 236–248; corrected data tables in *Appl. Stat.* **33,** 79–80.

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

HASEMAN, J. K. (1984). Statistical issues in the design, analysis, and interpretation of animal carcinogenicity studies. *Environ. Health Perspect.* **58,** 385–392.

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

KODELL, R. L., FARMER, J. H., GAYLOR, D. W., AND CAMERON, A. M. (1982). Influence of cause-of-death assignment on time-to-tumor analyses in animal carcinogenesis studies. *J. Natl. Cancer Inst.* **69,** 659–664.

LAGAKOS, S. W. (1982). An evaluation of some two-sample tests used to analyze animal carcinogenicity experiments. *Util. Math.* **21B,** 239–260.

LAGAKOS, S. W., AND LOUIS, T. A. (1985). The statistical analysis of rodent tumorigenicity experiments. In *Toxicological Risk Assessment* (Clayson, Krewski and Munro, eds.). CRC Press, Boca Raton, Fla.

MANTEL, N., AND HAENSZEL, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *J. Natl. Cancer Inst.* **22,** 719–748.

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

MITCHELL, T. J., AND TURNBULL, B. W. (1979). Log-linear models in the analysis of disease prevalence data from survival/sacrifice experiments. *Biometrics* **35,** 221–234.

National Toxicology Program (1982). Carcinogenesis bioassay of 1,2-dibromoethane in F344 rats and B6C3F1 mice (inhalation study). U.S. Dept. of Health and Human Services, Pub. Health Service, Natl. Inst. Health, Tech. Rep. No. 210.

PETO, R. (1974). Guidelines on the analysis of tumor rates and death rates in experimental animals (editorial). *Brit. J. Cancer* **29,** 101–105.

PETO, R., PIKE, M., DAY, N., GRAY, R., LEE, P., PARISH, S., PETO, J., RICHARDS, S., AND 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." International Agency for Research on Cancer Monographs, Supplement 2, 311–426.

RYAN, L. M. (1985). Efficiency of age-adjusted tests in animal carcinogenicity experiments. *Biometrics* **41,** 525–531.

TURNBULL, B. W., AND MITCHELL, T. J. (1978). Exploratory analysis of disease prevalence data from survival/sacrifice experiments. *Biometrics* **34,** 555–570.