

Sources of Variability in Rodent Carcinogenicity Studies

J. K. HASEMAN, J. E. HUFF, G. N. RAO, AND S. L. EUSTIS

National Institute of Environmental Health Science, Research Triangle Park, North Carolina 27709

Received August 10, 1988; accepted December 27, 1988

Sources of Variability in Rodent Carcinogenicity Studies. HASEMAN, J. E., HUFF, J. E., RAO, G. N., AND EUSTIS, S. L. (1989). *Fundam. Appl. Toxicol.* 12, 793-804. A number of factors may influence tumor rates in rodent carcinogenicity studies, including the animal room environment, genetic differences, food consumption/weight gain, survival/age of the animals, identification of gross lesions, pathology sampling procedures and preparation of the histology slides, and histopathologic diagnosis. The relative importance of these factors is evaluated, making use of laboratory animal carcinogenicity data from the National Toxicology Program and from other sources. An investigator must be aware of these potentially confounding factors, so that appropriate measures can be taken to reduce or eliminate their impact on the interpretation of study results. Certain potential sources of within-study variability can be controlled by appropriate experimental design and by proper conduct according to standard operating procedures. The effect of certain factors influencing tumor prevalence may be magnified when variability from study to study is considered, and thus it may be difficult to formulate a biologically meaningful statistical analysis that uses historical control data in a formal testing framework. © 1989 Society of Toxicology

Laboratory animal carcinogenicity studies, such as those carried out by the National Toxicology Program (NTP), the National Cancer Institute (NCI), and other national and international organizations, are utilized by the scientific community and by various government agencies in making regulatory decisions affecting public health. These experiments are important because, as noted by the International Agency for Research on Cancer (1987), "in the absence of adequate data on humans, it is biologically plausible and prudent to regard agents for which there is sufficient evidence of carcinogenicity in experimental animals as if they presented a carcinogenic risk to humans."

In the design, analysis, and interpretation of these studies, an investigator must maintain an awareness of the potentially confounding factors that may influence tumor prevalence both within and among experiments. This report summarizes and discusses

what we consider to be the major sources of variability in rodent carcinogenicity experiments, as determined by an evaluation of the NTP/NCI carcinogenicity study results as well as an examination of other large data sets.

POTENTIAL SOURCES OF VARIABILITY

A number of factors in addition to the chemical may influence tumor prevalence in long-term carcinogenicity studies. The more important of these are discussed below.

Animal Room Environment

Environmental variables that may influence the physiology of the animals and impact on toxicological findings include physi-

cal factors, such as light, temperature, relative humidity, ventilation, water, bedding, and diet, and biological factors, such as infections, diseases, and group compatibility of the animals in the experiments (Rao, 1986). Since many of these factors are comparable for control and chemically exposed animals within a given study (which are normally housed in the same room throughout the experiment), environmental variables are perhaps more important in determining among-study rather than within-study variability. However, even within a given study, a proper randomization of animals is essential to reduce or eliminate the effects of these variables on tumor prevalence.

In most studies animals are assigned at random to experimental groups, often after first stratifying by body weight. However, after this initial randomization, animals in some studies are housed by group in the same location in the animal room for the entire study period. While such housing simplifies maintenance of the animals, it is conceivable that certain factors associated with the animal room environment could selectively influence the biological responses of animals in certain cage locations, thus introducing a potential confounding factor into the analysis (Young, 1987, 1988; Haseman, 1988a,b).

For example, elevated rates of nonneoplastic eye lesions (cataracts and retinopathy) are frequently observed in rats or mice occupying the top row of cages in the rack (Greenman *et al.*, 1982; Rao, 1986; Haseman, 1988b). High light intensity is known to cause eye lesions in albino rodents (Rao, 1986), and animals in the top rows are nearer the fluorescent light source and thus exposed to more light than are animals in other cages.

In contrast, associations of cage location with tumor prevalence have not been well documented and are not consistent findings. Some investigators have reported that differences in tumor prevalence may be related to cage shelf level, but such associations have either been inconsistent (Lagakos and Motteler, 1981) or quite subtle, requiring thou-

sands of animals per group to attain statistical significance (Greenman *et al.*, 1984). Haseman *et al.* (1986) evaluated 18 long-term studies that utilized two concurrent control groups housed in separate locations in the animal room for the entire study, and found that the number of significant differences in tumor prevalence between the two control groups agreed closely with the number expected by chance. Thus, the animal room environment and caging protocols did not appear to contribute significantly to the observed tumor rates.

Conversely, Young (1987, 1988) reported that "local room effects" influenced the prevalence of hepatocellular neoplasms in one experimental group of male mice in a 2-year study of eugenol (NTP, 1983) and the prevalence of pancreatic acinar cell tumors in a control group of male rats in a carcinogenicity study of benzyl acetate (NTP, 1986). However, Haseman (1988a,b), in a broader evaluation of NTP studies with F344 rats and B6C3F1 mice, found that the frequency of such "significant effects" in these experiments was similar to chance expectation. Thus, it is likely that the apparent "local room effects" found by Young reflect random variability.

It is common practice to house laboratory animals three to five per cage for economic reasons and also because individual housing has been shown to cause stress in some instances (Hatch *et al.*, 1965; Sigg *et al.*, 1966). However, group housing introduces the possibility of "cage effects," i.e., the clustering of tumors within cages. This may impact on the selection of the proper experimental unit and the subsequent statistical analysis (Gart *et al.*, 1986). However, there has been little evidence to suggest that tumor prevalences cluster within cages, particularly when survival differences are taken into account (Gart, 1976; Arnold *et al.*, 1985; Gart *et al.*, 1986; Haseman, 1988a,b).

Because of problems associated with fighting among group-housed male mice, the NTP currently employs individual caging for

its studies with B6C3F1 mice. The NTP experimental design also includes random allocation of animals to cages, random assignment of columns of cages in a rack to chemically exposed or control groups, and periodic rotation of racks in the room and cage location in the racks (NTP, 1984).

Another environmental factor studied by the NTP is the possible influence of viral infection on tumor prevalence in Fischer 344 rats and B6C3F1 mice. Tumor rates were compared between NTP studies with viral infections (Sendai virus, pneumonia virus of mice, rat coronavirus/sialodacryoadenitis virus, mouse hepatitis virus) and studies without such infections. It was found that none of the viral infections appeared to affect the tumor rates in these studies when survival differences, interlaboratory variability, and time-related trends were taken into account (Rao *et al.*, 1989a,c).

Genetic Factors

Outbred stocks (e.g., ICR-Swiss, Sprague-Dawley) of rodents have low genetic stability marked by genetic drift between colonies (sources) and within the same colony over time. Even inbred strains from different sources separated by several generations may have subtle genetic differences. These differences coupled with environmental factors may influence toxicity (Walden and Schiller, 1985) and tumor prevalences (MacKenzie and Garner, 1973).

Within a given study it is conceivable that "litter effects" may influence tumor prevalence. Littermate information is generally not available when experimental animals are received from the supplier, but a proper randomization of animals to dosed and control groups should control this potential source of variability in any case.

An evaluation of possible litter effects in the ED01 study, which used approximately 24,000 animals, concluded that "In this study, one can reasonably assume that for the

occurrence of most tumor types, the prevalence rates do not vary significantly among litters" (Gaylor *et al.*, 1985). Other investigators have reached similar conclusions (Hase-man, 1988c). Specialized methods of statistical analysis have been proposed (Mantel *et al.*, 1977; Mantel and Ciminera, 1979; Gart *et al.*, 1986) if an investigator wishes to employ a litter-matched design. Also, for carcinogenicity studies in which animals are exposed *in utero*, the litter rather than the individual animal may be the appropriate experimental unit (Gart *et al.*, 1979).

Age and Survival of Animals

In some studies there may be apparent dose-related differences in tumor prevalence that simply reflect the greater survival of one group relative to the other. This underscores the need to take survival differences into account when assessing changes in tumor prevalence. Gart *et al.* (1986) provide an excellent discussion of the various survival-adjusted statistical methods that have been proposed for laboratory animal carcinogenicity studies.

Differences in the age of the animals may also influence tumor rates. For example, young animals (especially neonates) may be more susceptible to chemical-induced neoplasia than are older animals (Maltoni *et al.*, 1982; Drew *et al.*, 1983). All animals within a given study generally arrive in a single shipment and are thus approximately the same age. In contrast, animals from different studies may differ in age at the time of study onset, and the studies could be of varying durations (e.g., 18 months vs 24 months). Such differences may contribute to among-study variability in tumor rates.

Food Consumption/Weight Gain

Laboratory animals with lower body weights frequently have decreased tumor prevalence and improved survival relative to heavier animals (Ross and Bras, 1971; Con-

ybeare, 1980). Perhaps most striking is the association between lower body weights and the decreased prevalence of mammary gland neoplasms (see Haseman, 1983a; Rao *et al.*, 1987 and the references cited therein) and pituitary gland tumors (Ross *et al.*, 1970; Gries and Young, 1982; Rao *et al.*, 1987) in rats. There are also data to support a positive correlation between body weight and tumor incidence in humans (Doll and Peto, 1981).

This association may become an issue when animals exposed to chemicals consume less food (because of palatability, toxicity, or other factors) and gain less weight than concurrent controls. These chemically exposed animals often also show decreased tumor prevalence relative to controls. Thus, it is more difficult to detect chemically related carcinogenic effects in these animals, which may increase the false-negative rate.

Paired feeding is one possible solution to this problem, but such a study would be more time consuming and costly and might be difficult to implement effectively. Another approach would be to alter the common practice of *ad libitum* feeding and restrict food intake to a preselected amount for each animal. Alternatively, the feed could be made available only at preselected times and/or for limited periods. Modifications in the formulation of the diet should also be considered. Since 1980 the NTP has utilized the relatively high protein (24%) NIH 07 diet (Rao and Knapka, 1987), but is currently considering a reformulation with lower protein content, due in part to continuing concerns regarding possible dietary effects on tumor prevalence.

Tumor rates may be modulated by other dietary factors, including caloric restriction, amount and nature of fat in the diet, and amount of protein (Tannenbaum and Silverstone, 1952; Everett, 1984; Rao, 1986). We do not discuss these further except to point out the association of corn oil gavage with reduced rates of mononuclear cell leukemia and elevated prevalences of pancreatic acinar cell adenoma in male F344 rats (Haseman *et al.*, 1985). The mechanism by which corn oil

given by gavage affects these two tumor types is unknown (Eustis and Boorman, 1985).

Pathology

There are several individual pathology tasks beginning with the necropsy of the experimental animal and ending with the microscopic examination of the tissues that can contribute to the variability in reported tumor rates. These can be a source of potential bias within a study and more likely a source of variability among studies performed at different laboratories.

The histopathology sampling procedure in general use by the NTP is a combination of nonrandom sampling (all gross lesions observed at necropsy are examined microscopically) and random sampling (routine tissue sections are taken from all protocol required tissues). A thorough gross examination is important, since many of the tumors ultimately diagnosed microscopically are initially identified macroscopically by the prosector and pathologist. To minimize the possibility that tumors visible at gross necropsy will be overlooked, NTP procedures require that all tissues be saved following necropsy and be subject to an independent audit for untrimmed lesions.

Bias or variability can be introduced if the number of tissue samples taken for histopathology varies among groups. Under these conditions, a noncarcinogen might appear to be carcinogenic simply because of an increase in the number of sections evaluated and thus in the number of tumors detected (Ad Hoc Panel, 1984; Haseman, 1984). The potential for variability is reduced by utilizing protocols that specify set procedures for tissue sampling.

When slides are prepared, the amount of tissue trimmed, the site at which samples are taken, the orientation of the tissue on the slide, and the number of sections per animal should be similar in chemically exposed and control groups. For example, in one study the

histology technicians prepared cross sections of thyroid tissue for all control and most low-dose female rats, but prepared longitudinal sections for more than 60% of the high-dose female rats. The longitudinal sections included roughly twice the amount of thyroid tissue as the cross sections, and there was a concomitant increase in the prevalence of thyroid tumors in high-dose females relative to controls. This increase was attributed to the greater proportion of thyroid tissue examined microscopically in the high-dose group (NTP, 1988).

The histology technicians responsible for the study must be well trained and ideally should be assigned equal proportions of animals from each group to control this potential source of variability. Also, detailed standard operating procedures describing the histology functions are needed to prevent bias within a study and variability from study to study.

Histopathological examinations should follow stringent standards of quality, uniformity, and objectivity (Boorman and Eustis, 1986; Gart *et al.*, 1986). Because there may be variability among pathologists in histopathology nomenclature and diagnostic criteria, the same pathologist should examine all slides from chemically exposed and control groups in a species.

Further, even if a single pathologist is responsible for the histopathology of a given experiment, there may be time-related "diagnostic drifts" in classifying lesions. Since the histological criteria for distinguishing hyperplasia from neoplasia and benign from malignant tumors often consist of multiple factors of a subtle qualitative and/or quantitative nature, it is sometimes difficult to maintain consistency in the application of these criteria over time.

To avoid the possibility of diagnostic drift or subtle bias in identifying tumors, some investigators favor blinded pathology, i.e., histopathological examination without prior knowledge of whether the tissues are from chemically exposed or control animals. The

arguments for and against this practice have been debated (e.g., Fears and Schneiderman, 1973; Weinberger, 1980; Haseman, 1984; American College of Veterinary Pathologists, 1986). Perhaps a reasonable compromise between fully open and totally blind pathology is a procedure currently used by the NTP, which allows the study pathologist to diagnose tumors in a nonblind fashion, but then employs blind pathology during the pathology review phase of the study (Boorman and Eustis, 1986). Thus, all chemically related effects are verified under a more rigorous procedure that better ensures objectivity of diagnoses. Blind pathology is also important in the diagnosis of subtle toxic lesions.

The importance of histopathology diagnosis as a source of variability in tumor incidence is illustrated by considering the results of two independent examinations of the same slides from carcinogenicity studies of malathion and malaoxon (Reuber, 1985; Huff *et al.*, 1985). Remarkably different conclusions were reached by these two sets of investigators concerning tumor prevalence and the resulting carcinogenic potential of the chemicals studied. Examples of differences in histopathology diagnosis are given in Table 1 for male rats receiving malaoxon; similar differences were seen in the other experiments reported in these papers.

Fortunately, such major disagreements in tumor diagnosis do not appear to be commonplace. In NTP studies all tumors diagnosed and all target organs are subject to review by an independent pathologist (Boorman *et al.*, 1985), and differences of opinion are resolved by a pathology working group consisting of six to eight experts in rodent toxicologic pathology (Boorman and Eustis, 1986). This procedure reduces the likelihood that histopathology diagnosis will be a major source of within-study variability in tumor prevalence.

Random Differences

In laboratory animal carcinogenicity studies, as many as 30 to 40 different organ and

TABLE I

EXAMPLES OF DIFFERENCES OF HISTOPATHOLOGY DIAGNOSIS: TUMOR INCIDENCES FOR MALE F344 RATS IN THE MALAOXON STUDY AS DETERMINED BY REUBER (1985) AND BY HUFF *et al.* (1985)

Selected site-specific tumors	Reuber			Huff <i>et al.</i>		
	Control group	Low dose	High dose	Control group	Low dose	High dose
Adrenal gland						
Adenoma	4/50	9/50	11/46	0/50	0/50	0/49
Carcinoma	1/50	6/50	10/46	0/50	1/50	0/49
Leukemia	20/50	46/50	39/46	19/50	11/50	10/50
Liver						
Nodules	0/50	2/50	3/46	0/50	2/50	0/50
Carcinoma	0/50	2/50	6/46	1/50	0/50	1/50
Forestomach papilloma	1/50	12/50	12/46	0/47	1/48	0/48
All malignant tumors	35/50	49/50	42/46	28/50	16/50	26/50

tissue sites are examined for possible chemically related effects. In this situation statistically significant differences in tumor prevalence may occur by chance. It has been estimated that in a study involving male and female rats and mice, the probability is approximately 47–50% that some tumor type may show a statistically significant ($p < 0.05$) increased prevalence in a high-dose group relative to controls by chance (Haseman *et al.*, 1986). This underscores the need to be aware of the multiple comparisons issue, and not to regard every statistically significant increase in tumor prevalence automatically as a biologically meaningful effect. Biological as well as statistical factors must be considered in the overall evaluation of the data (Haseman, 1983b).

While the importance of random variability should not be minimized, the issue of false positives is well recognized by investigators in the field and generally taken into account in the overall evaluation of the data. For example, Gart *et al.* (1986) cite a number of studies that have demonstrated that "rules which attempt to model the actual decision process indicate that the false-positive rates are close to the nominal level." The Office of Science and Technology Policy (1985) reaches a sim-

ilar conclusion. Thus, the actual false-positive rates associated with laboratory animal carcinogenicity studies are much lower than the 47–50% figure given above, and have been estimated by Haseman (1983b) to be no greater than 7–8%. Moreover, a number of statistical procedures have been proposed to deal specifically with the false-positive issue (see, e.g., Fears *et al.*, 1977; Gart *et al.*, 1979; Brown and Fears, 1981; Haseman, 1983b; Heyse and Rom, 1987; Farrar and Crump, 1988).

IMPLICATIONS OF TUMOR VARIABILITY ON THE USE OF HISTORICAL CONTROL DATA

Although the concurrent control group is always the first and most appropriate control group used for decision making (Gart *et al.*, 1979; Tarone *et al.*, 1981), there are certain instances in which historical control information can aid in the overall evaluation of tumor prevalence. One example is the occurrence of rare or uncommon tumors, which may require less stringent statistical evidence of a chemical effect if a low background rate of the tumor can be documented from histor-

ical control data (Gart *et al.*, 1986). Another example is a marginally significant increase in tumor prevalence relative to concurrent controls, which may be discounted or considered to be biologically meaningful when historical control data are considered (Tarone *et al.*, 1981; Society of Toxicology, 1982). Tabulations of historical control tumor prevalences from NTP carcinogenicity studies have been published (e.g., Solleveld *et al.*, 1984; Haseman *et al.*, 1984, 1985), and historical control rates for tumors showing evidence of chemically related effects also appear in NTP Technical Reports.

A number of statistical procedures have been proposed for incorporating historical control data in a formal testing framework, and a good discussion of these techniques is given by Krewski *et al.* (1988). However, there is currently no consensus regarding how and when historical control data should be used in the decision-making process (Krewski *et al.*, 1988; Haseman, 1988c). Some investigators have been critical of these procedures on statistical grounds (Tamura and Young, 1986), but there are also biological factors that may limit the use of historical control data. If historical control data are to be used in a meaningful way, the experiments in the data base must be similar to the current experiment in factors known to affect tumor rates (Gart *et al.*, 1986). This is often difficult, because each experiment is unique, and the effect of the factors discussed above may be magnified when variability in tumor prevalences from study to study is considered.

A number of investigators have reported significant variability in control tumor rates among laboratories (Tarone *et al.*, 1981; Haseman *et al.*, 1984, 1986). For example, Haseman *et al.* (1986) examined tumor rates from a series of 18 carcinogenicity studies conducted in rats and mice. The majority of these studies (13/18) were carried out in two laboratories—6 in Laboratory A and 7 in Laboratory B. Although the age of the animals at terminal sacrifice varied among the 13 studies, the average length of study (in

days; mean \pm SE) was similar at Laboratory A (856 ± 23) and Laboratory B (850 ± 21).

The observed rates of adrenal gland cortical tumors in the dual-control groups of female rats from these experiments are given in Table 2. Within a study, there was little difference in tumor rates between the two concurrent control groups. Laboratory A maintained relatively consistent tumor rates across studies, while there was considerable variability among the studies evaluated at Laboratory B. Moreover, the overall prevalence of adrenal gland cortical tumors in control female rats at Laboratory B (280/968, 29%) was approximately 15 times the rate observed at Laboratory A (17/839, 2%).

The prevalences of certain tumors also appear to be increasing over time (Everett, 1984; Haseman *et al.*, 1984; Rao *et al.*, 1989b). For example, the rates of hematopoietic system tumors (primarily mononuclear cell leukemia) in male F344 rats averaged less than 10% in control animals from 2-year NCI studies conducted during 1971–1974 (Fig. 1). During 1980–1983 the rate was almost 50%. Less striking increases were observed in other commonly occurring tumors in both male and female F344 rats (Rao *et al.*, 1989b).

To determine whether or not changes in pathology diagnosis might be contributing to the apparent changing tumor rates over time, the NTP initiated a pathology reevaluation of slides from untreated control animals from representative early and later studies for those tumors showing the most pronounced time-related differences. The results of this evaluation for male and female rats are summarized in Table 3 (taken from Rao *et al.*, 1989b).

This pathology reevaluation resulted in consistently higher rates of leukemia, thyroid gland C-cell tumors, and adrenal gland pheochromocytomas and lower incidences of anterior pituitary gland tumors than originally diagnosed. While changes in diagnostic criteria may have contributed to the apparent time-related increases in leukemia, adrenal gland pheochromocytoma, and pituitary gland neoplasms, the increasing time-related

TABLE 2
INCIDENCE OF ADRENAL CORTICAL TUMORS IN DUAL CONTROL GROUPS OF FEMALE RATS
FROM 13 COLOR ADDITIVE STUDIES

Compound	Laboratory A control rates ^a	Compound	Laboratory B control rates ^a
FD&C Blue 1	1/70-0/70	FD&C Blue 2	16/69-17/69
FD&C Yellow 5	2/69-0/70	FD&C Green 3	37/69-39/70
FD&C Red 3	1/70-0/70	FD&C Yellow 6	37/70-34/69
D&C Red 6	2/70-1/70	D&C Yellow 10	9/69-3/70
D&C Red 21	3/70-1/70	D&C Red 19	7/69-7/69
D&C Red 33	2/70-4/70	D&C Orange 5	32/69-32/69
		D&C Orange 17	5/68-5/69
Overall rate	17/839 (2%)	Overall rate	280/968 (29%)

^a Tumor incidences observed in each of two separate "identical" concurrent control groups.

trends were evident for most tumors, based on both the original and new diagnoses (see Table 3).

Changes in the amount of tissue examined may also have contributed to the time-related differences in the rates of pituitary gland, thyroid gland, and adrenal gland medullary tumors. For example, the median anterior pituitary gland tissue size for male rats was only 5.0 mm² in the 1972-1973 studies compared with 9.0 mm² in the 1980-1981 studies, and if these two sets of studies are matched by

amount of tissue examined, there is no longer a significant difference in tumor prevalence between them. A similar result holds for adrenal gland and thyroid gland tumors (Rao *et al.*, 1989b).

However, this finding is difficult to interpret because the presence of a tumor in a tissue will likely result in a larger section for examination. Thus, it is not clear if more tumors were detected because of the increased amount of tissue examined, or alternatively, if the increased prevalence of tumors resulted in larger sections for examination.

Rats in the more recent studies had higher body weights, which may also have contributed to the increased tumor rates observed in these animals (Rao *et al.*, 1989b).

In summary, there are a number of factors that may produce variability in tumor rates from study to study. Thus, even if a universally accepted statistical analysis could be developed for utilizing historical control data in a formal testing framework, such analyses may not be appropriate on biological grounds because of the lack of comparability of animals in the contemporary study to those from earlier studies in the data base.

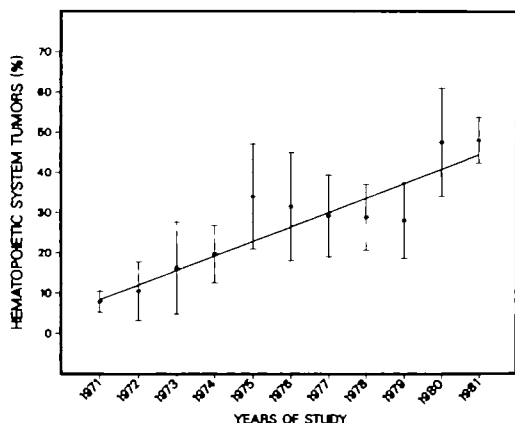


FIG. 1. Incidence of hematopoietic system tumors in untreated male F344 rats in NCI/NTP studies (mean \pm SD). Least-squares regression line is given. Tumor incidence shows significant ($p < 0.001$) linear increase over time.

CONCLUSIONS

What are the practical implications of these potential sources of variability? First, within-study variability should be minimized by a

TABLE 3

ORIGINAL AND REEVALUATED PREVALENCES FOR LEUKEMIA, ANTERIOR PITUITARY TUMORS, THYROID C-CELL TUMORS, AND ADRENAL PHEOCHROMOCYTOMAS IN UNTREATED MALE AND FEMALE F344 RATS FOR FOUR SELECTED TIME PERIODS

Year study starts	No. of animals	Leukemia (%)		Anterior pituitary tumors (%)		Thyroid C-cell tumors (%)		Adrenal pheochromocytomas (%)	
		O	R	O	R	O	R	O	R
Male									
1972-1973	244	13	23	11	9	9	11	11	27
1975-1976	250	33	44	18	17	10	14	18	29
1978-1979	250	30	41	26	23	11	17	23	26
1980-1981	249	44	50	26	26	15	20	31	38
Female									
1972-1973	247	11	21	37	28	7	10	^a	^a
1975-1976	248	21	30	44	29	9	14	^a	^a
1978-1979	250	16	28	50	27	8	18	^a	^a
1980-1981	248	27	31	49	34	19	23	^a	^a

Note. O = Original tumor rates; R = reevaluated tumor rates; data taken from Rao *et al.* (1989b).

^a Tumors not reevaluated.

proper experimental design. Specific recommendations include appropriate randomization of animals; control of environmental factors; consistency of gross necropsy, slide preparation, and histopathology diagnosis across groups; and comprehensive quality assurance and pathology review procedures.

Second, in the statistical analyses of tumor data, appropriate measures should be taken to minimize the impact of potential confounding factors that are likely to influence the overall interpretation of the study. One such example is the utilization of survival-adjusted methods to take into account differences in age at death between chemically exposed and control groups.

Third, the false-positive and false-negative issues must be considered when interpreting experimental results. Because of the multiplicity of sites examined, differences in tumor rates among groups may occur by chance. While the statistical significance of an observed effect is important in the overall evaluation of the data, biological factors must always be considered.

Fourth, because of the importance of the various factors that influence tumor rates across studies, care should be exercised in using historical control data in the overall evaluation of rodent carcinogenicity studies. The most relevant historical data are from contemporary studies at the same laboratory. Even in this instance, however, it is unclear whether or not historical data should be utilized in a formal testing framework.

These results also suggest that efforts to construct large data bases of tumor rates obtained by pooling data from studies at different laboratories carried out at different times under different experimental protocols may be of limited scientific value if the intent of such a data base is to provide a reference point for comparing tumor prevalence in a particular study. There are simply too many uncontrolled sources of variability in tumor rates. To have any value, such efforts should include as a minimum a reexamination of slides to ensure some consistency of diagnosis across studies.

Finally, one should not overinterpret the implications of variability in tumor rates. Us-

ing the concurrent control group as the primary basis for comparison eliminates the impact of among-study variability. Further, proper utilization of scientific principles in the design, conduct, analysis, and interpretation of laboratory animal studies will minimize the impact of many sources of within-study variability. Thus, the mere existence of variability in response does not invalidate an experiment. The critical issue is to maintain an awareness of potential confounding factors that may affect tumor prevalence, so that appropriate measures can be taken to reduce or eliminate their influence on the design, analysis, and interpretation of the study.

REFERENCES

- Ad Hoc Panel (1984). *Report of the NTP ad Hoc Panel on Chemical Carcinogenesis Testing and Evaluation*. NTP Board of Scientific Counselors, August 17, 1984.
- American College of Veterinary Pathologists (1986). Letter to the Editor. *Toxicol Appl Pharmacol* **83**, 184-185.
- ARNOLD, D. L., MOODIE, C. A., CHARBONNEAU, S. M., GRICE, H. C., MCGUIRE, P. F., BRYCE, F. R., COLLINS, B. T., ZAWIDZKA, Z. Z., KREWSKI, D. R., NERA, E. A., AND MUNRO, I. C. (1985). Long term toxicity of hexachlorobenzene in the rat and the effect of dietary vitamin A. *Food Chem Toxicol* **23**, 779-793.
- BOORMAN, G. A., AND EUSTIS, S. L. (1986). The Pathology Working Group as a means for assuring pathology quality in toxicological studies. In *Managing Conduct and Data Quality of Toxicology Studies*. (B. K. Hoover, J. K. Baldwin, A. F. Uelner, C. E. Whitmire, C. L. Davies, and D. W. Bristol, Eds.), pp. 271-275. Princeton Scientific, Princeton, NJ.
- BOORMAN, G. A., MONTGOMERY, C. A., JR., EUSTIS, S. L., WOLFE, M. J., MCCONNELL, E. E., AND HARDISTY, J. F. (1985). Quality assurance in pathology for rodent carcinogenicity studies. In *Handbook of Carcinogen Testing* (H. A. Milman and E. K. Weisburger, Eds.), pp. 345-357. Noyes, Park Ridge, NJ.
- BROWN, C. C., AND FEARS, T. R. (1981). Exact significance levels for multiple binomial testing with application to carcinogenicity screens. *Biometrics* **37**, 763-774.
- CONYBEARE, G. (1980). Effect of quality and quantity of diet on survival and tumor incidence in outbred Swiss mice. *Food Cosmet. Toxicol* **18**, 65-75.
- DOLL, R., AND PETO, R. (1981). The causes of cancer. Quantitative estimates of avoidable risks of cancer in the United States today. *J Natl Cancer Inst* **66**, 1191-1308.
- DREW, R. T., BOORMAN, G. A., HASEMAN, J. K., MCCONNELL, E. E., BUSEY, W. M., AND MOORE, J. A. (1983). The effect of age and exposure duration on cancer induction by a known carcinogen in rats, mice, and hamsters. *Toxicol Appl Pharmacol* **68**, 120-130.
- EUSTIS, S. L., AND BOORMAN, G. A. (1985). Proliferative lesions of the exocrine pancreas: Relationship to corn oil gavage in the National Toxicology Program. *J Natl. Cancer Inst* **75**, 1067-1073.
- EVERETT, R. (1984). Factors affecting the spontaneous tumor incidence rates in mice: A literature review. *CRC Crit. Rev Toxicol* **13**, 105-113.
- FARRAR, D., AND CRUMP, K. (1988). Exact statistical tests for any carcinogenic effect in animal bioassays. *Fundam Appl Toxicol* **11**, 652-663.
- FEARS, T. R., AND SCHNEIDERMAN, M. A. (1973). Pathologic evaluation and the blind technique (Letter to the Editor). *Science* **183**, 1144.
- FEARS, T. R., TARONE, R. E., AND CHU, K. C. (1977). False-positive and false-negative rates for carcinogenicity screens. *Cancer Res* **37**, 1941-1945.
- GART, J. J. (1976). Statistical analysis of the first mouse skin painting study. In *Report No 1 Toward Less Hazardous Cigarettes* (G. B. Gor, Ed.), DHEW Publication No. (NIH) 76-905, pp. 110-148. National Institutes of Health, U.S. Government Printing Office, Washington, DC.
- 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.
- GART, J. J., KREWSKI, D., LEE, P. N., TARONE, R. E., AND WAHRENDORF, J. (1986). *Statistical Methods in Cancer Research*. Vol. III. *The Design and Analysis of Long-Term Animal Experiments*. IARC Scientific Publication No. 79. International Agency for Research on Cancer, Lyon.
- GAYLOR, D. W., CHEN, J. J., GREENMAN, D. L., AND THOMPSON, C. H. (1985). Occurrence of tumors among litters of BALB/c female mice. *J Natl Cancer Inst* **74**, 803-809.
- GREENMAN, D. L., BRYANT, P., KODELL, R. L., AND SHELDON, W. (1982). Influence of cage shelf level on retinal atrophy in mice. *Lab. Anim. Sci* **32**, 353-356.
- GREENMAN, D. L., KODELL, R. L., AND SHELDON, W. B. (1984). Association between cage shelf level and spontaneous and induced neoplasms in mice. *J Natl Cancer Inst* **73**, 107-113.
- GRIES, C. L., AND YOUNG, S. S. (1982). Positive correlation of body weight with pituitary incidence in rats. *Fundam Appl. Toxicol* **2**, 145-148.
- HASEMAN, J. K. (1983a). Patterns of tumor incidence in two-year cancer bioassay feeding studies in Fischer 344 rats. *Fundam Appl Toxicol* **3**, 1-9.

- HASEMAN, J. K. (1983b). A reexamination of false-positive rates for carcinogenicity studies. *Fundam. Appl. Toxicol.* **3**, 334-339.
- HASEMAN, J. K. (1984). Statistical issues in the design, analysis, and interpretation of animal carcinogenicity studies. *Environ. Health Perspect.* **58**, 385-392.
- HASEMAN, J. K. (1988a). Lack of cage effects on liver tumor incidence in B6C3F1 mice. *Fundam Appl Toxicol.* **10**, 179-187.
- HASEMAN, J. K. (1988b). Do cage effects influence tumor incidence? An examination of laboratory animal carcinogenicity studies utilizing Fischer 344 rats. *J. Appl. Toxicol.* **8**(4), 267-273.
- HASEMAN, J. K. (1988c). Value of historical control data. In *Carcinogenicity: The Design, Analysis, and Interpretation of Long-Term Animal Studies*, pp. 53-54. International Life Sciences Institute, Springer-Verlag, New York.
- HASEMAN, J. K., HUFF, J., AND BOORMAN, G. A. (1984). Use of historical control data in carcinogenicity studies in rodents. *Toxicol. Pathol.* **12**, 126-135.
- HASEMAN, J. K., HUFF, J. E., RAO, G. N., ARNOLD, J. E., BOORMAN, G. A., AND MCCONNELL, E. E. (1985). Neoplasms observed in untreated and corn oil gavage control groups of F344/N rats and (C57BL/6N × C3H/HeN)F1 (B6C3F1) mice. *J. Natl. Cancer Inst.* **75**, 975-984.
- HASEMAN, J. K., WINBUSH, J. S., AND O'DONNELL, M. W., JR. (1986). Use of dual control groups to estimate false positive rates in laboratory animal carcinogenicity studies. *Fundam. Appl. Toxicol.* **7**, 573-584.
- HATCH, A., WIBERG, G., ZAWIDZKA, Z., CANN, M., AIRTH, J., AND GRICE, H. (1965). Isolation syndrome in the rat. *Toxicol. Appl. Pharmacol.* **7**, 737-745.
- HEYSE, J. F., AND ROM, D. (1987). Adjusting for multiplicity of statistical tests in the analysis of carcinogenicity studies using multivariate randomization tests. Paper presented at the 147th Annual Meeting of the American Statistical Association, San Francisco, August 17, 1987.
- HUFF, J. E., BATES, R., EUSTIS, S. L., HASEMAN, J. K., AND MCCONNELL, E. E. (1985). Malathion and malaoxon: Histopathological reexamination of the National Cancer Institute's carcinogenesis studies. *Environ. Res.* **37**, 154-173.
- International Agency for Research on Cancer (1987). IARC Monographs on the Evaluation of Carcinogenic Risks to Humans—Preamble. IARC Internal Technical Report 87/001, IARC, Lyon.
- KREWSKI, D., SMYTHE, R. T., DEWANJI, A., AND COLIN, D. (1988). Statistical tests with historical controls. In *Carcinogenicity: The Design, Analysis, and Interpretation of Long-Term Animal Studies*, pp. 23-38. International Life Sciences Institute, Springer-Verlag, New York.
- LAGAKOS, S., AND MOSTELLER, F. (1981). A case study of statistics in the regulatory process: The FD&C Red 40 experiments. *J. Natl. Cancer Inst.* **66**, 197-212.
- MACKENZIE, W. F., AND GARNER, F. M. (1973). Comparison of neoplasms in six sources of rats. *J. Natl. Cancer Inst.* **50**, 1243-1257.
- MALTONI, C., LEFEMINE, G., CILIBERTI, A., COTTI, G., AND CARETTI, D. (1982). Vinyl chloride: A model carcinogen for risk assessment. In *Genotoxic Effects of Airborne Agents*, eds. (R. R. Tice, D. L. Costa, and K. M. Schaich, Eds.), pp. 329-344. Plenum, New York.
- MANTEL, N., BOHIDAR, N. R., AND CIMINERA, J. L. (1977). Mantel-Haenszel analyses of litter-matched time-to-response data, with modifications for recovery of interlitter information. *Cancer Res.* **37**, 3863-3868.
- MANTEL, N., AND CIMINERA, J. L. (1979). Use of log rank scores in the analysis of litter-matched data on time to tumor appearance. *Cancer Res.* **39**, 4308-4315.
- National Toxicology Program (1983). *Carcinogenesis Studies of Eugenol in F344/N Rats and B6C3F1 Mice (Feed Studies)*, NTP Technical Report 223. U.S. Department of Health and Human Services, National Toxicology Program, Research Triangle Park, NC.
- National Toxicology Program (1984). *General Statement of Work for the Conduct of Acute, 14-Day Repeated Dose, 90-Day Subchronic, and 2-Year Chronic Studies in Laboratory Animals*. National Toxicology Program, Research, Triangle Park, NC.
- National Toxicology Program (1986). *Toxicology and Carcinogenesis Studies of Benzyl Acetate in F344/N Rats and B6C3F1 Mice (Gavage Studies)*, NTP Technical Report 250. U.S. Department of Health and Human Services, National Toxicology Program, Research Triangle Park, NC.
- National Toxicology Program (1988). *Toxicology and Carcinogenesis Studies of Penicillin VK in F344/N Rats and B6C3F1 Mice (Gavage Studies)*, NTP Technical Report 336. U.S. Department of Health and Human Services, National Toxicology Program, Research Triangle Park, NC.
- Office of Science and Technology Policy (1985). Chemical carcinogens: A review of the science and its associated principles. *Fed. Regist.*, March 14, 1985, Part II, 10371-10442.
- RAO, G. N. (1986). Significance of environmental factors on the test system. In *Managing Conduct and Data Quality of Toxicology Studies* (B. K. Hoover, J. K. Baldwin, A. F. Uelner, C. E. Whitmire, C. L. Davies, and D. W. Bristol, Eds.), pp. 173-186. Princeton Scientific, Princeton, NJ.
- RAO, G. N., HASEMAN, J. K., AND EDMONDSON, J. (1989a). Influence of viral infections on body weight, survival, and tumor prevalence in Fischer 344 rats of long-term studies. *Lab. Anim. Sci.*, in press.
- RAO, G. N., HASEMAN, J. K., GRUMBAIN, S., CRAWFORD, D. D., AND EUSTIS, S. L. (1989b). Growth, body

- weight, survival, and tumor trends in F344/N rats during an eleven year period. Submitted for publication.
- RAO, G. N., AND KNAPKA, J. J. (1987). Contaminant and nutrient concentrations of natural ingredient rat and mouse diet used in chemical toxicity studies. *Fundam Appl. Toxicol.* **9**, 329-338.
- RAO, G. N., PIEGORSCH, W. W., CRAWFORD, D. D., EDMONDSON, J., AND HASEMAN, J. K. (1989c). Influence of viral infections on body weight, survival and tumor prevalences of B6C3F1 (C57Bl/6N × C3H/HeN) mice in carcinogenicity studies. *Fundam. Appl. Toxicol.*, in press.
- RAO, G. N., PIEGORSCH, W. W., AND HASEMAN, J. K. (1987). Influence of body weight on the incidence of spontaneous neoplasms in rats and mice of long-term studies. *Amer. J. Clin. Nutr.* **45**, 252-260.
- REUBER, M. B. (1985). Carcinogenicity and toxicity of malathion and malaoxon. *Environ. Res.* **37**, 119-153.
- ROSS, M. H., AND BRAS, G. (1971). Lasting influence of early caloric restriction in prevalence of neoplasms in the rat. *J. Natl. Cancer Inst.* **47**, 1095-1113.
- ROSS, M. H., BRAS, G., AND RAGBEER, M. S. (1970). Influence of protein and caloric intake upon spontaneous tumor incidence of the anterior pituitary gland of the rat. *J. Nutr.* **100**, 177-189.
- SIGG, E., DAY, C., AND COLOMBO, C. (1966). Endocrine factors in isolation-induced aggressiveness in rodents. *Endocrinology* **78**, 679-684.
- Society of Toxicology (1982). Animal data in hazard evaluation: Paths and pitfalls. *Fundam Appl. Toxicol.* **2**, 101-107.
- SOLLEVELD, H. A., HASEMAN, J. K., AND MCCONNELL, E. E. (1984). Natural history of body weight gain, survival, and neoplasia in the F344 rat. *J. Natl. Cancer Inst.* **72**, 929-940.
- TAMURA, R. N., AND YOUNG, S. S. (1986). The incorporation of historical control information in tests of proportions: Simulation study of Tarone's procedure. *Biometrics* **42**, 343-349.
- TANNENBAUM, A., AND SILVERSTONE, H. (1952). The genesis and growth of tumors. V. Effects of varying the level of B vitamins in the diet. *Cancer Res.* **12**, 744-749.
- TARONE, R. E., CHU, K. C., AND WARD, J. M. (1981). Variability in the rate of some common naturally occurring tumors in Fischer 344 rats and (C57Bl/6N × C3H/HeN)F1 (B6C3F1) mice. *J. Natl. Cancer Inst.* **66**, 1175-1181.
- WALDEN, R., AND SCHILLER, C. M. (1985). Comparative toxicity of 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD) in four (sub) strains of adult male rats. *Toxicol. Appl. Pharmacol.* **77**, 490-495.
- WEINBERGER, M. A. (1980). How valuable is blind pathology in histopathologic examinations in conjunction with animal toxicity studies? *Toxicol. Pathol.* **7**, 14-17.
- YOUNG, S. S. (1987). Are there local room effects on hepatic tumors in male mice? An examination of the NTP Eugenol study. *Fundam Appl. Toxicol.* **8**, 1-4.
- YOUNG, S. S. (1988). Evaluation of data from long-term rodent studies (Letter to the Editor). *J. Natl. Cancer Inst.* **80**, 3-4.