WEINBERGER, M. A. (1979). How valuable is blind evaluation in histopathological examinations with animal toxicity studies? *Toxicol. Pathol.* **7** 14–17.

WEINBREN, K. (1978). The liver. In *Systematic Pathology* **3**. Churchill Livingstone, London.

WEISBURGER, E. K. (1981). *N*-substituted aryl compounds in carcinogenesis and mutagenesis. *NCI Monograph* **58**.

WEISBURGER, E. K. (1983). Species differences in response to aromatic amines. In *Organs and Species Specificity in Chemical Carcinogenesis* (R. Langenbach, S. Nesnow and J. M. Rice, eds.). Plenum, New York.

WHITTEMORE, A. S., GROSSER, S. C. and SILVERS, A. (1986). Pharmacokinetics in low dose extrapolation using animal cancer data. *Fundam. Appl. Toxicol.* **7** 183–190.

WHITTEMORE, A. and KELLER, J. (1978). Quantitative theories of carcinogenesis. *SIAM Rev.* **20** 1–30.

WILBOURN, J. D., ET AL. (1986). Response of experimental animals to human carcinogens: An analysis based upon the IARC Monographs programme. *Carcinogenesis* **7** 1853–1863.

WILLIAMS, G. M. (1980). Classification of genotoxic and epigenetic hepatocarcinogens using liver culture assays. *Ann. N. Y. Acad. Sci.* **349** 273–282.

WILLIAMS, G. M. (1983). Epigenetic effect of liver tumor promoters and implications for health effects. *Environ. Health Perspect.* **50** 177–183.

WILSON, R. and CROUCH, E. A. C. (1987). Risk assessment and comparisons: An introduction. *Science* **236** 267–270.

ZEISE, L., WILSON, R. and CROUCH, E. (1984). Use of acute toxicity to estimate carcinogenic risk. *J. Risk Anal.* **4** 187–99.

Comment: Risk Assessment: Science or Policy?

Norman Breslow

Freedman and Zeisel present a rather devastating critique of quantitative risk assessment as practiced by certain government agencies, largely by demonstrating the lack of sound scientific evidence to support the assumptions that enter into the process. Along the way they attack the multistage model of carcinogenesis as having no basis in biology, question the logic of low-dose extrapolation, cast doubt on the relevance of animal experiments for evaluating potential humans risks and summarize the evidence for one particular chemical with the statement "... taking the results of the bioassays at face value, DDT seems on balance to inhibit tumor development."

Few scientists would dispute their claim that current procedures used for routine risk assessment on the basis of limited animal data lack a solid scientific foundation. Freedman and Zeisel apparently would urge us to abandon quantitative risk assessment altogether until scientific advances permit the construction of "realistic" statistical models of the underlying biological processes. I believe that this goal is illusory, that they are much too pessimistic about the contributions science can make to the regulatory process and that in their zeal to argue their point of view they have made scientific errors as serious as those in the work they criticize.

Norman Breslow is Professor and Chairman, Department of Biostatistics, University of Washington, Seattle, Washington 98195.

THE NATURE OF RISK ASSESSMENT

Freedman and Zeisel seem to imply early on in their paper that risk assessments are an exercise in science and, starting from this premise, they proceed to demolish the scientific argument. Only in the penultimate Section 7 do we learn that risk assessments generally are not viewed by their protagonists as science in the usual sense, but rather involve decision making in the face of uncertainty. The point is well put in the report of the Committee on the Institutional Means for Assessment of Risks to Public Health (National Academy of Sciences, 1983a, page 11):

> "The dominant and analytic difficulty is pervasive uncertainty. Risk assessment draws extensively on science, and a strong scientific basis has developed for linking exposure to chemicals to chronic health effects. However, data may be incomplete, and there is often great uncertainty in estimates of the types, probability, and magnitude of health effects associated with a chemical agent, of the economic effects of a proposed regulatory action, and of the extent of current and possible future human exposures."

And further (page 12):

> "Ethical considerations prevent deliberate human experimentation with potentially dangerous chemicals, and the length of the latent period for cancer and some other effects greatly complicates epidemiologic studies of uncontrolled human exposures. Animal models must be used to investigate whether exposure to a chemical is related to the incidence of health effects, and the results must be extrapolated to humans. To make judgments amid such uncertainty, risk assessors must rely on a series of assumptions."

The assumptions that the Committee refers to, which include such items as selection of a model for extrapolation, they later term "inference guidelines" in order to recognize explicitly that they involve value judgments and policy choices in areas where scientific knowledge is limited. (See also the 1982 document from the State of California Department of Health Services, Carcinogen Identification Policy: A Statement of Science as a Basis of Policy.)

The current trend toward quantitative risk assessment needs to be placed in historical perspective against a background of advances in analytic chemistry that made feasible the detection of trace amounts of toxic chemicals in human tissue, foodstuffs and ambient air and water. Given the pervasive public fear of cancer and birth defects, the alternatives to quantitative risk assessment in some cases may well be a total ban on otherwise valuable chemical products.

THE MULTISTAGE MODEL

Freedman and Zeisel review the Armitage-Doll multistage model that underlies some of the dose-time-response models fitted to both animal and human data and find that it departs considerably from scientific reality. They distinguish the biological version of the model, in which a "colony of cells progresses through stages on the way to cancer," from the statistical version in which an "individual cell executes a Markov chain through a fixed order of states . . .," lamenting the statisticians' failure to identify the states with particular biological phenomena. Their discussion of the biological model is largely limited to the classical initiation-promotion paradigm developed by experimentalists working with whole animals. Much of the current evidence supporting the model, however, comes from molecular genetics. A good recent statement is that of Barrett (1987): "The underlying premise of this model is that cancer is a multistep process and that the cancer cell evolves by clonal expansion of a cell which acquires sequentially different cellular properties." Two of the steps, identified on the basis of both in vivo and in vitro studies, involve inactivation of tumor suppressor gene(s) and activation of a "transforming" oncogene. Barrett notes that the steps need not occur in order and that there are likely to be multiple pathways to the same cancer end point.

The modern molecular viewpoint of cancer arising from a single cell that has undergone a sequence of heritable changes and then formed a cancer clone broadly supports the line of reasoning that led to the original formulation of the multistage model. Of course, at any point in time there may be many cells in intermediate stages of carcinogenesis. The rates at which such cells divide and multiply, which could be quite different from the rates for "normal" tissue, will influence the time or age of tumor occurrence (Moolgavkar and Knudson, 1981).

In one situation I am familiar with, we have a pretty good grasp on the specific cellular events leading to tumor formation. The story of how we came to this discovery is both interesting and pertinent. Some 15 years ago, Knudson and Strong (1972) compiled statistical data from case reports on the distribution of age at diagnosis for children with Wilms' tumor (a rare childhood kidney tumor). From these simple observations and their earlier work with retinoblastoma, they postulated a two-stage mutational model for the origin of Wilms' tumors. Cases in which the first mutation was prezygotic constituted the "hereditary" type, thought to include the familial and bilateral cases, whereas those in which it was postzygotic constituted the "sporadic" cases. At the time, of course, they had no idea what these mutagenic events might be. A few years later it was discovered that children with both Wilms' tumor and aniridia invariably had a constitutional deletion of chromosomal material at the 11p13 locus (Francke et al., 1979). Recent molecular studies of Wilms' tumor patients using DNA probes have suggested that an abnormal somatic segregation results in a cell that is homozygous or hemizygous at this same locus (Koufos et al., 1984). Thus, the molecular evidence is rather strong for a cancer process that initially involves a mutation (possibly inactivation of a tumor suppressor gene) followed by loss of the wild type allele via chromosomal reduplication or nondysfunction. Although some of the epidemiological inferences made by Knudson and Strong were shaky, so that they overestimated the fraction of "heritable" tumors (Breslow and Beckwith, 1982), the two-step mutational process that they hypothesized now has been confirmed in the laboratory. Clearly, one need not wait for the specific molecular events that underlie an observed statistical regularity to be discovered before developing a mathematical model for the process that has useful predictive properties.

Mendelian genetics perhaps provides the best known example of this type, but there are others.

Bearing this point in mind, we might recall that the Armitage-Doll model was developed before the era of modern molecular biology in order to explain observed regularities in the age incidence patterns for epithelial tumors. Contrary to the assertions of Freedman and Zeisel, my experience suggests that the age-specific incidence rates of such tumors are in fact remarkably consistent with the power law between the ages of 35 and 65 or so for many different human populations. Departures from this basic pattern, which most often take the form of a downturn at older ages, can often be attributed to a birth cohort phenomenon or to problems of under diagnosis and under reporting at older ages. Armed with new biological insights, we now need to see if we can refine the model to account for other statistical regularities. The work of Moolgavkar and colleagues is a step in this direction.

Although at one time, according to Freedman and Zeisel, the "multistage model seemed like a promising avenue to explore," they imply that it has not stood up to scientific scrutiny and evidently would have us abandon it. I would argue, on the contrary, that Armitage and Doll were on the right track and that we should continue the search for models both to explain the available epidemiological data and to guide the biologists in their research.

LOW-DOSE EXTRAPOLATION

I have more sympathy with Freedman and Zeisel's views on the vagaries of low-dose extrapolation. The apparent desire of some regulatory agencies to modify the design of the routine NTP bioassay, in order to enhance its power to discriminate between risk assessment models that yield quite different estimates of "virtually safe doses," is very disturbing to me. Knowing the shape (within the usual limits of statistical uncertainty) of the dose-response function in the dose range at which one observes tumors in laboratory animals is of little help in predicting the shape at the minute doses of usual concern to the risk assessor. A preferable strategy would be to base the policy decision involved in selection of the risk model on scientific judgements about probable mechanisms, taking into account any available information on metabolic pathways, detoxification, chemical disposition, the capacity of the agent to react with human DNA in vitro, etc. Statistical data from studies of a few hundred animals treated at moderate to high doses is perhaps best used to identify potential carcinogens and to rank order them in terms of potency, as in estimation of the TD_{50} (Sawyer et al., 1984).

Regarding the policy decision itself, however, I believe the scientific community has an obligation to consider the problem and to share its insights with those in positions of authority. A reasonable concensus has been achieved that low-dose linear extrapolation is appropriate for setting standards for exposure to ionizing radiation (BEIR, 1980), even though the human data on low-dose exposures that would fully justify this approach do not presently exist and are unlikely to be forthcoming (Gilbert, 1985). Of course, one has a substantial body of radiobiological theory and mountains of data from both animal and human studies to guide the process. Chemicals are much more difficult because of the need for metabolic activation and transport to the site of action, because the relevant animal and human data are generally much more limited and because there are so many of them to contend with. This does not mean we should give up!

Freedman and Zeisel fail to mention one of the more cogent arguments in favor of using linear extrapolation to predict increases in human cancer incidence (above background levels) that may accompany carcinogenic exposure at low doses. If the toxic agent produces tumors via one of the pathways responsible for the spontaneous cases, it may be viewed as acting to increment the "dose" of whatever "naturally" occurring agents are responsible for those cases. This concept, known as dose-wise additivity (Peto, 1978), is implicit in the usual multistage model formulation of an agent that increases transition rates between stages. The Abbots' formula used by Freedman and Zeisel in their example, however, implicitly assumes that the toxic agent acts quite independently of the background causes (Tomatis et al., 1980).

PREDICTIVE VALUE OF THE ANIMAL BIOASSAY

Freedman and Zeisel's comments on "the man to mouse argument" and "consistency with epidemiology" are particularly specious. In their Tables 6 and 7 they seem to imply that a chemical classified as having "limited" or "inadequate" evidence of carcinogenicity should be regarded as somehow less carcinogenic than one for which there is "sufficient" evidence of carcinogenicity. This is a blatant distortion of the IARC definitions. According to IARC (1982)

> "Limited evidence of carcinogenicity, . . . means that the data suggest a carcinogenic effect but are limited because: (a) the studies involve a single species, strain, or experiment; or (b) the experiments are restricted by inadequate dosage levels, inadequate duration of exposure to the agent, inadequate period of follow-up, poor survival, too few animals, or inadequate reporting; or (c) the neoplasms produced often occur spontaneously

and, in the past, have been difficult to classify as malignant by histological criteria alone (e.g., lung and liver tumors in mice)."

The degree of evidence for carcinogenesis tells us nothing at all about carcinogenic potency. Highly carcinogenic compounds may have "limited" evidence of carcinogenicity because the major experiment on which the evaluation was based involved a single species or single dose level, so that no opportunity existed for demonstrating a dose-response trend. Arsenic, a well known human carcinogen, was for years listed as having "inadequate" evidence for carcinogenicity in the rodent bioassay because the doses needed to produce a toxic effect killed most of the animals before the ages at which tumors generally arose.

What one needs in order to evaluate the sensitivity, specificity and predictive value of the bioassay is a data base of chemicals that have been *adequately tested* for carcinogenicity in both animals and humans. The appropriate two-way table will have axes labeled "positive" versus "negative" or "strong" versus "weak" according to the results of such tests. The fact that such a data base does not exist reflects two realities. First, an adequate test means treating a large enough group of subjects (mice or men) with a sufficiently high dose (but not so high that all the subjects die young) so that a clear response will be observed if the agent is carcinogenic. For humans, this essentially means limiting the study to common substances in everyday use (tobacco, alcohol, vitamin C, sugar, medications) or to industrial agents where there has been moderately heavy workplace exposure (arsenic, asbestos). Adequate epidemiology is impossible for most chemicals. For animals, it means that some company or government agency was sufficiently concerned about the chemical's carcinogenicity to want to spend several hundreds of thousands of dollars to evaluate it. There is an obvious selection bias so that chemicals which, because of chemical structure or mutagenicity in short term tests, stand a good chance of testing positive in the bioassay are the very ones selected for it.

A similar problem affects the argument for discordance among rodent species. We are told that "65%, 58% and 23%" of 26 human carcinogens were found to be carcinogenic in the mouse, the rat and the hamster, respectively. Although we thus know the numerator data, the number of chemicals that tested positive in each bioassay, we are not provided with the denominators, the number for which an adequate test was carried out. I urge the reader of this series to examine the paper by Wilbourn et al. (1986), which updates the "data" in Freedman and Zeisel's Tables 6 and 7, before coming to any conclusion about the predictive value of animal bioassays for human carcinogenesis.

REVIEW OF THE DATA ON DDT

Freedman and Zeisel present data in Table 9 on Z tests for dose-response trends in site-specific tumor incidence according to the dose-level of DDT exposure. Because this was compiled from published sources there was no real opportunity to correct the data for the evident fact that "DDT shortens the lifespan." Unfortunately, this deficiency means that the data presented are virtually uninterpretable in terms of the relation between DDT exposure and cancer occurring at a particular site. The reason is evident from Table 1 and Figure 1, which are reproduced from a study involving over 4000 mice exposed to DDT and urethane, a known rodent carcinogen (Breslow et al., 1974). (This study is an extension of the Tomatis 1972 study in their Table 9; the data were also reported by

TABLE 1
Cumulative percentages of animals surviving by treatment group, sex and age

Treatment	Sex	No. of Animals	Percentage alive at				
			0 wk	70 wk	90 wk	110 wk	130 wk
Control	M	348	100	80.7	63.5	32.5	13.8
	F	363	100	79.1	60.0	35.0	14.0
DDT (2 ppm)	M	362	100	83.7	57.2	22.4	3.0
	F	354	100	77.4	55.1	23.7	5.6
DDT (10 ppm)	M	367	100	81.7	57.2	25.1	8.2
	F	370	100	83.8	64.9	42.4	15.4
DDT (50 ppm)	M	396	100	84.6	52.8	21.5	3.5
	F	349	100	80.2	56.4	26.9	6.3
DDT (250 ppm)	M	372	100	72.3	28.5	1.3	0.0
	F	334	100	66.5	41.3	9.9	1.5
Urethane	M	315	100	72.1	39.7	8.9	0.1
	F	248	100	72.2	40.7	12.1	1.2

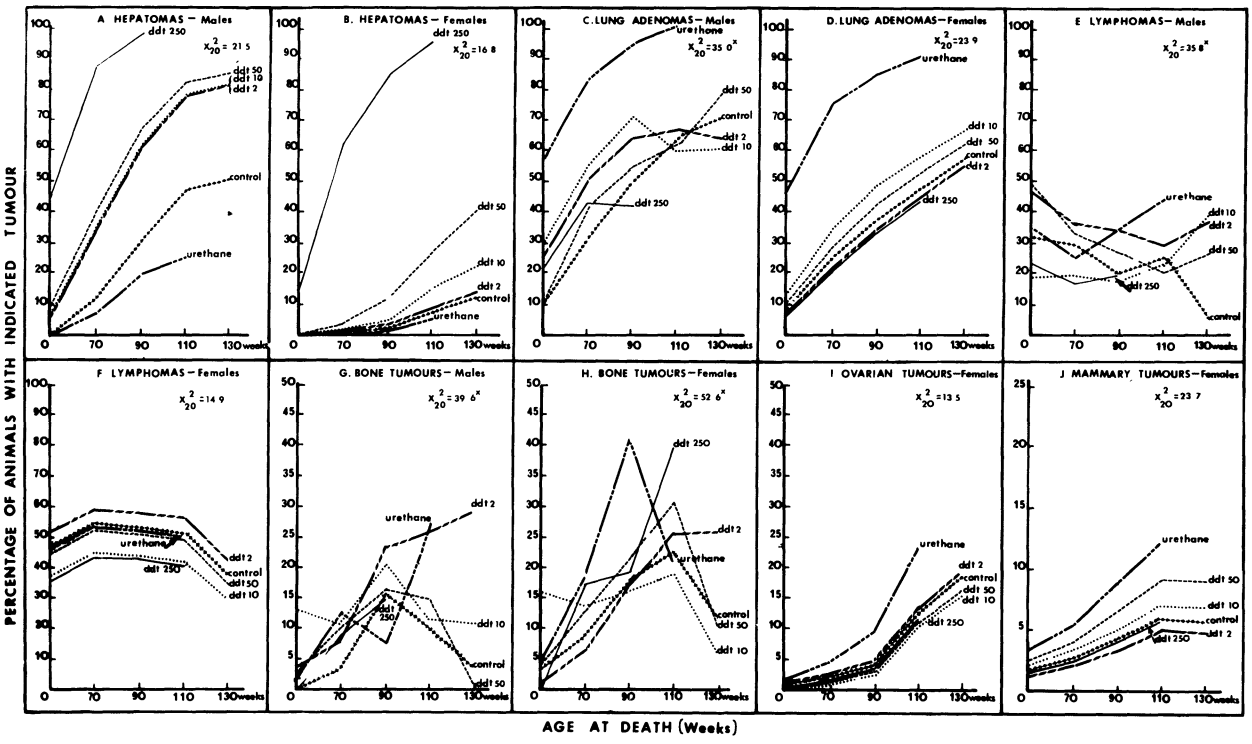

FIG. 1. *Age prevalence curves. Proportions of animals with indicated tumor type by sex, treatment group and age at necropsy.* χ^2 *indicates the fit of the linear logistic model.*

Turusov et al., 1973.) Regrettably, about half the panels of Figure 1, those for which χ^2 is not significant, present not original data but data smoothed by a linear logistic model. Nevertheless, the message should be clear.

DDT indeed severely curtailed the life-span of treated animals. At 250 ppm exposure, fewer than 50% survived beyond 90 wk. Lower exposures, with the exception of females treated at 10 ppm, affected mortality to a lesser but still noticeable extent. On the other hand, lung tumors were most likely to occur in animals brought to necropsy at 90 wk of age or after. When examined on an age-specific basis (Figure 1), there is little if any evidence for a negative trend in lung tumor prevalence with dose of DDT. Adjusting for age as a single continuous covariate in a logistic regression model for tumor prevalence (Dinse and Lagakos, 1983), I calculate $Z = -1.7$ for males and $Z = -0.7$ for females. Without adjustment for age, on the other hand, there is a strong negative trend ($Z = -5.0$ for males and $Z = -4.0$ for females) because of the deficit of lung tumors in mice treated at 250 ppm, most of whom died before they could develop them. Even if, using the version of these data compiled in Table 3 of Turusov et al. (1973), we make the crude age adjustment suggested by Freedman and Zeisel, using as denominators the "effective number" of animals alive at the time of appearance of the first tumor, we find $Z = -4.6$ for males and $Z = -3.4$ for females. No wonder that in Freedman and Zeisel's Table 9, 6 out of the 7 experiments with a Z value of less than -2.0 for lung give evidence for a positive effect of DDT on overall mortality. A similar phenomenon could well explain the negative trend observed for other sites (ovaries, mammary gland, kidneys) where there is a marked increase in tumor prevalence with age at necropsy.

The prevalence of mouse lymphoma does not depend strongly on age (Figure 1). Here the negative Z's noticed by Freedman and Zeisel are more intriguing. In our study of 4000 mice, we observed a negative association between liver tumors and lymphomas at the level of the individual animal that could not readily be explained by age at death, treatment or a postulated competing risks phenomenon. Could it be that mice susceptible to hepatoma are resistant to lymphoma, or that the two carcinogenic processes somehow interfere with one another? We hope that the experimentalists will give some further thought to this provocative observation that has now been replicated in grouped data (Haseman, 1983).

In any event, without appropriate age adjustment, the admittedly tongue-in-cheek claim of Freedman and Zeisel that DDT inhibits overall tumor formation is completely indefensible.

CONCLUSIONS

In summary, I find that Freedman and Zeisel have misconstrued the nature of risk assessment, ignored

important molecular evidence that supports the concept of multistage carcinogenesis, presented a questionable method (Abbots' formula) to account for background tumor incidence, distorted the IARC definitions of sufficient and limited evidence and failed to adjust for a major confounding factor (age) in their analysis of the DDT data. In spite of all this, I share in large measure their skepticism about the scientific value of routine risk assessments that use statistical models fitted to limited animal data obtained at high doses to predict the human response at low ones. Society needs critics like Freedman and Zeisel to challenge establishment viewpoints, lest the repeated use of "inference guidelines" such as low-dose linear extrapolation lends them undeserved credence. Hopefully, other scientists will continue their constructive efforts to improve the biological and statistical models and to contribute their expertise to the decision making process.

ADDITIONAL REFERENCES

BARRETT, J. C. (ed.) (1987). A multistep model for neoplastic development: Role of genetic and epigenetic changes. In *Mechanisms of Environmental Carcinogenesis II: Multistep Models of Carcinogenesis*. CRC Press, Boca Raton, Fla.

BEIR (1980). *The Effects on Populations of Exposure to Low Levels of Ionizing Radiation*. Report of the Advisory Committee on the Biological Effects of Ionizing Radiations, National Academy of Sciences, Washington.

BRESLOW, N. E. and BECKWITH, J. B. (1982). Epidemiological features of Wilms' tumor: Results of the National Wilms' Tumor Study. *J. Nat. Cancer Inst.* **68** 429–436.

BRESLOW, N. E., ET AL. (1974). Associations between tumor types in a large-scale carcinogenesis study of CF-1 mice. *J. Nat. Cancer Inst.* **52** 233–239.

CALIFORNIA STATE DEPARTMENT OF HEALTH SERVICES (1982). Carcinogen identification of policy: A statement of science as a basic of policy.

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

FRANCKE, V., ET AL. (1979). Aniridia-Wilms' tumor association: Evidence for specific deletion of 11p13. *Cytogenet. Cell. Genet.* **24** 185–192.

GILBERT, E. S. (1985). How much can be learned from populations exposed to low levels of radiation? *Statistician* **34** 19–30.

KNUDSON, A. G. and STRONG, L. C. (1972). Mutation and cancer: A model for Wilms' tumor of the kidney. *J. Nat. Cancer Inst.* **48** 313–324.

KOUFOS, A., ET AL. (1984). Loss of alleles at loci on human chromosomes during genesis of Wilms' tumor. *Nature* **309** 170–172.

PETO, R. (1978). Carcinogenic effects of exposure to very low levels of carcinogenic substances. *Environ. Health Perspect.* **22** 155–159.

Comment

J. K. Haseman

I was disappointed to find Freedman and Zeisel taking such a one-sided and negative position concerning the scientific value of laboratory animal studies for assessing possible cancer risks to humans. The scientific merits of laboratory animal studies and quantitative risk estimation have been debated for years, and Freedman and Zeisel raise no new points that have not been considered extensively elsewhere. The difference between their article and more definitive publications (e.g., Office of Science and Technology Policy, 1985) is that Freedman and Zeisel make no effort to present a balanced view on the major issues.

Freedman and Zeisel utilize several questionable techniques to achieve their objectives. These include (1) selectively citing references that appear to support their point of view while ignoring other publications that express contrary views, and (2) misrepresenting or misinterpretating data from various sources, in some cases reaching the opposite conclusion to that given by the original investigator. Examples of this will be given throughout these comments, which will be limited primarily to the area of qualitative risk assessment.

Throughout the paper Freedman and Zeisel display an arrogant attitude toward nonstatisticians, (e.g., assuming that investigators do not randomize properly unless the randomization scheme is stated explicitly; claiming that "pathologists see themselves as professionals exempt from bias"). This air of superiority, especially when considering biological issues, reduces their own credibility and the credibility of all statisticians in the eyes of biologists, many of whom feel that statisticians and lawyers debating science is no more meaningful than biologists debating p-values.

The major criticisms of laboratory animal carcinogenicity studies cited by Freedman and Zeisel include the following.

J. K. Haseman is Research Mathematical Statistician, Biometry and Risk Assessment Program, National Institute of Environmental Health Sciences, P. O. Box 12233, Research Triangle Park, North Carolina 27709.