Risk Analysis, Vol. 10, No. 4, 1990

Comment

The Importance of Ranking Possible Carcinogenic Hazards Using HERP[1]

Lois Swirsky Gold,[2] Leslie Bernstein[3], and Bruce N. Ames[2]

Testing chemicals for carcinogenicity at near-toxic doses in rodents does not provide enough information to predict the excess number of human cancers that might occur at low-dose exposures. It is better to admit this than to provide the public with worst-case scenarios or to pretend that QRA is scientifically justifiable. The HERP index uses the same animal results and similar statistical methods as the usual low-dose linear estimation of risk; however, our purpose is to *compare* possible carcinogenic hazards from a variety of naturally occuring and synthetic chemicals, not to perform risk assessments. Our ranking is based on a simple measure that indicates what percentage of a standardized rodent tumorigenic dose a human gets from a given exposure. Recently, we have discussed advances in understanding the role of cell proliferation in the mechanisms of carcinogenesis, which suggest that estimates of risk to humans from low doses of rodent carcinogens have been markedly overestimated. The high doses administered in animal cancer tests are postulated to induce chronic cell proliferation, which itself is mutagenic in several ways.[1,2] Since cell proliferation due to toxicity is not observed at low doses, the cancer risk at low doses is likely to be much lower than previously thought, particularly for nongenotoxic compounds. As more theory is developed and more evidence is produced about the mechanisms of carcinogenesis, the ranking of hazards by the simple HERP index can be improved (as can risk assessment) by taking into account information on a given chemical about mechanism, shape of the dose response, and mutagenicity.

About half of the chemicals tested for carcinogenicity in rats and mice are positive, and this proportion is similar for naturally occurring and synthetic chemicals.[1] Because of the mechanistic considerations discussed above, we expect that a high percentage of all chemicals would be positive if tested at maximally tolerated doses (MTD). Thus it becomes important to try to rank possible carcinogenic hazards to humans from exposures to various chemicals. We have argued that the whole natural world has been neglected in the field of testing chemicals for carcinogenicity.[1,3,4] Yet, humans are exposed to millions of natural chemicals. For example, less than 0.1% of exposures to chemicals in the diet (by weight) are likely to be from synthetic chemicals. Because there are so many natural chemicals, they should be used as a reference for evaluating carcinogenic hazards from synthetic chemicals. Therefore, the emphasis in our work is upon comparing and ranking human exposures in order to achieve some perspective on the natural background of carcinogens and to suggest priorities for epidemiological investigation. A chemical pollutant should not be a high priority for concern with respect to carcinogenicity if its possible hazard seems far below that of many common food items. To evaluate possible hazard we determine how close a typical human exposure is to the estimated dose that will have the probability of an animal remaining tumor free by the end of a standard lifetime (TD_{50}). The TD_{50} is calculated using a linear model and requires little or no extrapolation from the bioassay doses. We have based the HERP analysis on our standardized database of animal cancer tests, the Carcinogenic Potency Database (CPDB), which currently includes results of 3969 experiments on 1052 chemicals.[5–8]

Our ranking of possible carcinogenic hazards suggests that human exposures to rodent carcinogens are common in everyday life, and that the possible hazards of synthetic chemicals ingested from pesticide residues or water pollution appear to be trivial relative to the background of rodent carcinogens from natural and traditional chemicals (e.g., from the cooking of food). Our

[1] Received January 5, 1990.

[2] Lawrence Berkeley Laboratory, 1 Cyclotron Road, Berkeley, California 94720.

[3] Department of Preventive Medicine, University of Southern California School of Medicine, Los Angeles, California 90033.

0272-4332/90/1200-0625$06.00/1

results also suggest that alcohol at moderate doses should be high on our priority list for epidemiological studies and cancer prevention, and that possible hazards in the workplace may be of particular concern.[9,10] In general, we would expect similar results from a ranking of risk estimates derived from QRA, since the methods are similar to those used to calculate TD_{50}.

Wartenberg and Gallo (W&G) have attempted to show by a counterexample of two experiments that the HERP ranking is inaccurate because the TD_{50} does not provide the correct rank order of carcinogenic potencies at low exposures (within the range of doses tested) when the dose–response in one test is quadratic. They compare one TD_{50} from a 62-week mouse test of AF-2 to one from a 135-week rat test of vinyl chloride. Their conclusion is incorrect because they apparently did not realize that TD_{50} is a dose-rate for a *standard lifespan* (2 years in rodents). The apparent rank reversal of TD_{50}'s in their example is due solely to their not taking into account our adjustment for differences in experiment length. Before extrapolation to the standard lifetime, the order of TD_{50}'s is what they consider correct (i.e., vinyl chloride is more potent than AF-2 for these two cases). When an experiment is terminated before the standard lifespan, animals are not at risk of developing tumors later in life. Thus the number of tumors found will be reduced, and the TD_{50} will be greater than the true TD_{50} (i.e., the compound will appear to be less potent than it actually is). Because tumor incidence increases markedly with time, the reverse is true for unusually long experiments. Our convention in the calculation of TD_{50} has been to adopt as a correction factor f^2, where f is experiment time/standard lifespan.[11] Generally, in the CPDB, extrapolation is minimal because of the inclusion rules and the use of standard protocols. Only 6% of the mouse tests are as short as 62 weeks, and only 6% of rat tests are as long as 135 weeks.

Figures 1 and 2 indicate the experimental results (solid line) on which TD_{50} is based, as well as the values for TD_{50}, before extrapolation to a 2-year lifespan (broken line), and 99% confidence limits (dotted line) for the AF-2 and vinyl chloride data referred to by W & G. The unextrapolated TD_{50} for vinyl chloride (506 mg/kg/day) is more potent than the unextrapolated TD_{50} for AF-2 (640 mg/kg/day). The unextrapolated confidence limits overlap [vinyl chloride (317,1160), and AF-2 (398,1120)]. Since the AF-2 test ended at 62 weeks and the vinyl chloride test ended at 135 weeks, the extrapolation to 2 years alters the ranking of TD_{50} for these two cases. As reported in the CPDB, when TD_{50} is extrapolated to a standard lifespan, the value for vinyl chloride becomes less potent (843 mg/kg/day) than that for AF-2 (225 mg/kg/day). After the time of extrapolation, the confidence limits do not overlap [vinyl chloride (528,1940) and

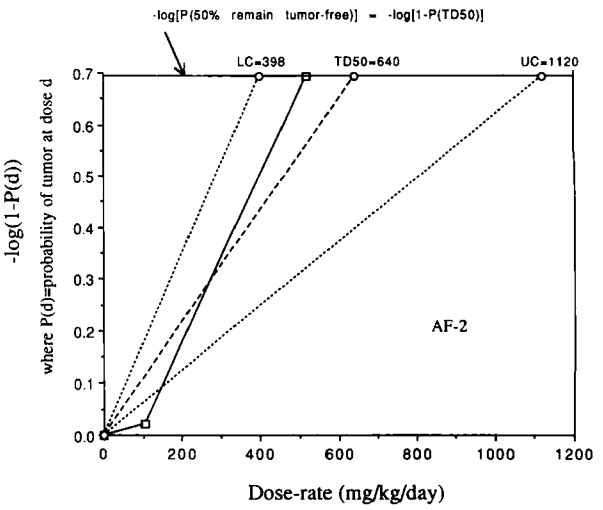

Fig. 1. Unextrapolated TD_{50} and unextrapolated confidence limits from a 62 week experiment of AF-2 discussed by Wartenberg and Gallo. Solid line is for experimental results for squamous-cell carcinomas of the forestomach in female ICR mice. LC = lower confidence limit. UC = upper confidence limit.

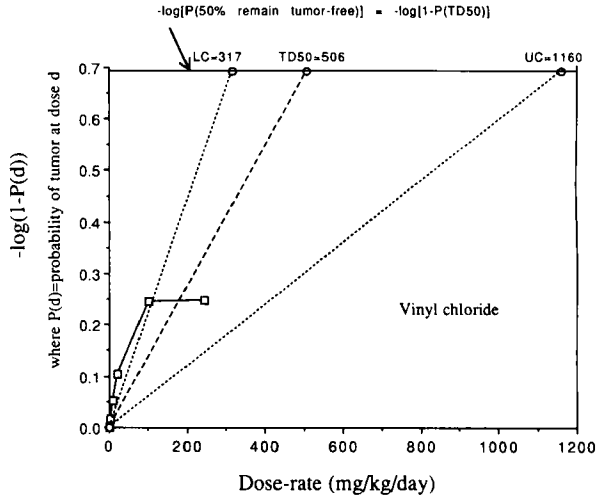

Fig. 2. Unextrapolated TD_{50} and unextrapolated confidence limits from a 135 week experiment of vinyl chloride discussed by Wartenberg and Gallo. Solid line is for experimental results for liver angiosarcomas in both sexes of Sprague-Dawley rats. LC = lower confidence limit. UC = upper confidence limit.

AF-2 (140,393)]. The large impact of the standardization to a 2-year lifespan reflects the unusual experiment lengths of these two tests and the fact that tumor yields increase markedly with time.

Additionally, neither of the TD_{50} values used by W&G would enter into the HERP calculation. The TD_{50} used in HERP is the average calculated by taking the harmonic mean of the TD_{50}'s of the positive tests in a species; for each test the target site with the lowest TD_{50} value is used. The harmonic mean is calculated separately for rats and mice, and the more sensitive species is used in the HERP. W&G selected target sites that were not the most potent in each experiment. For example, in the case of AF-2 in mice they selected the incidence for squamous-cell carcinomas of the forestomach (0/65, 1/50, 25/50). The most potent target site (the lowest TD_{50} value) in that experiment is for the mix of squamous-cell papillomas and squamous-cell carcinomas of the forestomach (0/65, 13/50, 36/50), which reflects the morphologic continuum from papillomas to carcinomas and the tumor types that might be expected by 62 weeks. This TD_{50} is consistent with linearity, and is the value that would be used in the calculation of the harmonic mean for mice. The CPDB also includes results for rats on AF-2, and the harmonic mean of the rat TD_{50}'s is lower than that for mice.[5] Thus, the value used in the HERP for TD_{50} is for rats, whereas W&G have based their analysis on a test in mice.

There are additional unusual aspects to the cases chosen by W&G. As noted above, the AF-2 test was unusually short and the vinyl chloride test unusually long. Vinyl chloride is an unusual chemical because of the plateau in the dose–response. The standard protocol in carcinogenesis testing is a control group and two dose groups at the MTD and half the MTD. In comparison, the vinyl chloride test has a control and six dose groups, and the doses range over 200-fold. Fewer than 2% of the positive chemicals in rats vary that much in their design. For AF-2 in mice, the two doses range over five fold; only 11% of the mouse carcinogens have been tested over such a wide dose range.

While it is theoretically possible for a rank reversal in carcinogenic potency to occur, W&G have failed to demonstrate it with their example. We would not expect such reversals to be common. In earlier work we showed that given the usual experimental design and the experimental result that 100% of the animals rarely develop the tumor of interest in an experiment, the TD_{50} calculated for a statistically significant result will be close to the administered dose. Moreover, if a quadratic model were used to estimate potency, TD_{50} would not vary greatly from the estimate based on a linear model. We showed for the standard protocol of the MTD and half the MTD that "the (exponential) linear assumption for the dose–response function used in calculating TD_{50} is not crucial. For example, assuming a purely (exponential) quadratic dose–response function, we found the actual TD_{50} to be within a factor of 5 of the linearly estimated TD_{50} for a selection of background rates and incidence rates at the *max-d* [maximum dose tested]."[12]

In addition, we have shown that various methods of summarizing TD_{50} across experiments provide estimates that deviate only over an extremely small range. We considered most potent site, harmonic mean (of the most potent site from each positive experiment), geometric mean, and arithmetic mean.[13] Only a few chemicals are extreme with respect to the variation in TD_{50} values, and AF-2 and vinyl chloride are two of them, as we reported earlier.[13]

REFERENCES

1. B. N. Ames, and L. S. Gold, "Perspective: Too Many Rodent Carcinogens: Mitogenesis Increases Mutagenesis." *Science* **249**, 970–971 (1990).
2. B. N. Ames, and L. S. Gold, "Chemical Carcinogenesis: Too Many Rodent Carcinogens." *Proc. Natl. Acad. Sci.* **87**, 7772–7776 (October 1990).
3. B. N. Ames, M. Profet, and L. S. Gold, "Dietary Pesticides (99.99% All Natural)." *Proc. Natl. Acad. Sci.* **87**, 7777–7781 (October 1990).
4. B. N. Ames, M. Profet, and L. S. Gold, "Nature's Chemicals and Synthetic Chemicals: Comparative Toxicology." *Proc. Natl. Acad. Sci.* **87**, 7782–7786 (October 1990).
5. L. S. Gold, C. B. Sawyer, R. Magaw, G. M. Backman, M. de Veciana, R. Levinson, N. K. Hooper, W. R. Havender, L. Bernstein, R. Peto, M. C. Pike, and B. N. Ames, "A Carincogenic Potency Database of the Standardized Results of Animal Bioassays." *Environ. Health Perspect.* **58**, 9–319 (1984).
6. L. S. Gold, M. de Veciana, G. M. Backman, R. Magaw, P. Lopipero, M. Smith, M. Blumenthal, R. Levinson, L. Bernstein, and B. N. Ames, "Chronological Supplement to the Carcinogenic Potency Database: Standardized Results of Animal Bioassays Published Through December 1982." *Environ. Health Perspect.* **67**, 161–200 (1986).
7. L. S. Gold, T. H. Slone, G. M. Backman, R. Magaw, M. Da Costa, P. Lopipero, M. Blumenthal, and B. N. Ames, "Second Chronological Supplement to the Carcinogenic Potency Database: Standardized Results of Animal Bioassays Published Through December 1984 and by the National Toxicology Program Through May 1986." *Environ. Health Perspect.* **74**, 237–329 (1987).
8. L. S. Gold, T. H. Slone, G. M. Backman, S. Eisenberg, M. Da Costa, M. Wong, N. B. Manley, L. Rohrbach, and B. N. Ames, "Third Chronological Supplement to the Carcinogenic Potency Database: Standardized Results of Animal Bioassays Published Through December 1986 and by the National Toxicology Program Through June 1987." *Environ. Health Perspect.* **84**, 215–286 (1990).
9. L. S. Gold, G. M. Backman, K. Hooper, and R. Peto, "Ranking the Potential Carcinogenic Hazards to Workers from Exposures to

Chemicals That Are Tumorigenic in Rodents."*Environ. Health Perspect.* **76**, 211–219 (1987).
10. B. N. Ames, R. Magaw, and L. S. Gold, "Ranking Possible Carcinogenic Hazards." *Science* **236**, 271–280 (1987).
11. R. Peto, M. C. Pike, L. Bernstein, L. S. Gold, and B. N. Ames, "The TD_{50}: A Proposed General Convention for the Numerical Description of the Carcinogenic Potency of Chemicals in Chronic-Exposure Animal Experiments." *Environ. Health Perspect.* **58**, 1–8 (1984).
12. L. Bernstein, L. S. Gold, B. N. Ames, M. C. Pike, and D. G. Hoel, "Some Tautologous Aspects of the Comparison of Carinogenic Potency in Rats and Mice." *Fund. Appl. Toxicol.* **5**, 79–86 (1985).
13. L. S. Gold, T. H. Slone, and L. Bernstein, "Summary of Carcinogenic Potency (TD_{50}) and Positivity for 492 Rodent Carcinogens in the Carcinogenic Potency Database." *Environ. Health Perspect.* **79**, 259–272 (1989).

Risk Analysis, Vol. 10, No. 4, 1990

The Fallacy of Ranking Possible Carcinogen Hazards Using the TD_{50}

Daniel Wartenberg[1,2] and Michael A. Gallo[1]

Received June 6, 1988; revised May 14, 1990

Ames *et al.* have proposed a new model for evaluating carcinogenic hazards in the environment. They advocate ranking possible carcinogens on the basis of the TD_{50}, the estimated dose at which 50% of the test animals would get tumors, and extrapolating that ranking to all other doses. We argue that implicit in this methodology is a simplistic and inappropriate statistical model. All carcinogens are assumed to act similarly and to have dose–response curves of the same shape that differ only in the value of one parameter. We show by counterexample that the rank order of cancer potencies for two chemicals can change over a reasonable range of doses. Ames *et al.*'s use of these TD_{50} ranks to compare the hazards from low level exposures to contaminants in our food and environment is wholly inappropriate and inaccurate. Their dismissal of public health concern for environmental exposures, in general, based on these comparisons, is not supported by the data.

KEY WORDS: Risk assessment; TD_{50}; HERP; carcinogens; human health.

1. INTRODUCTION

Quantitative risk assessment (QRA) is the process by which scientists predict adverse health outcomes in low-exposure situations using data derived from high-exposure situations (e.g., rodent bioassays, human worker studies). Beginning with simple statistical models for susceptibility distributions,[1–4] investigators have tried to improve QRA methodology by including a variety of more complicated biologically based statistical distribution models,[5–7] time-to-tumor models,[8–11] models that incorporate background tumor rates[12–13] and, most recently, models that incorporate mechanisms of biological activity in conjunction with statistics.[14–17] Thus, it was with much disappointment that we read a recent article by Ames *et al.* (Ref. 18; see also 19 and 20) that newly proposed a simple, purely statistical and inappropriate model for assessing human health hazard. Their model relies only on the rank order of cancer potencies derived from exposures at arbitrary, high levels, and yet is claimed to be appropriate for consideration of human health hazard at all exposure levels. Our concern lies not in their proposal itself, but in the widespread notoriety it has gained[21] and its acceptance by risk assessment practitioners in the field.[3] We show below that the formulation of Ames *et al.* is wholly inappropriate for the assessment of human hazard. Ames *et al.*[18–20] claim that the absolute risk estimates from QRA are not sufficiently accurate, based on heuristic considerations. Based on theoretical considerations, we contend that using only their relative potencies, one cannot even find the correct rank order of carcinogenic potencies at low exposures.

2. THE AMES *ET AL.* MODEL

In their recent articles, Ames *et al.*[18–20] suggested that, in consideration of potential human health risk, carcinogens should be ranked on the basis of their TD_{50},

[1] Department of Environmental and Community Medicine, Robert Wood Johnson Medical School, University of Medicine and Dentistry of New Jersey, Piscataway, New Jersey.

[2] To whom all correspondence should be addressed.

[3] For example, USDA Forest Service, Pacific Northwest Region, *Managing Competing and Unwanted Vegetation*, Draft Environmental Impact Statement (Portland, Oregon, October 1987).

0272-4332/90/1200-0609$06.00/1

the estimated dose which would elicit a median response among test animals.[4] Combining this index with actual human exposures to these same chemicals, they calculated an alleged human hazard, the HERP. We argue that this approach to hazard evaluation is fallacious on two accounts. First, human (or animal) response to different chemicals may not follow the same dose–response function which is an inherent characteristic of the TD_{50} evaluation. To the degree that these functions differ for different chemicals, the rank order of potencies at the TD_{50} may not correspond to the rank order of potencies at typical environmental exposures (10^{-5} or 10^{-6} quantal response). Second, even if the hazards do have similar functional forms, all the parameters used to describe the specific functions save one are held constant. That is, if one considers the likelihood of response as a statistical distribution (as with the probit model described below), implicit in the TD_{50} concept is the equality of variances *in response* to different chemicals. Only the central tendency of the function can vary. Not only is the constancy of all parameters but one in responses to different chemicals statistically unlikely, this assumption precludes a hazard evaluation which could incorporate variability among different biological or metabolic responses to different chemicals. Moreover, we believe that adoption of the TD_{50} as an index of that dose which would produce a specific effect in humans (potency) will counteract the recent advances in toxicological modeling and, once again, relegate dose–response evaluation to purely a statistical exercise. Acceptance of our argument implies that the ranking of carcinogens provided by Ames *et al.*[18–20] is not necessarily related to human hazard. We do not address statistical estimation issues in this note because if the index is flawed, as we contend, there is no need to estimate its values.[22]

3. THE TD_{50} CONCEPT

The basic conceptual model behind the TD_{50} index is that the rank order of cancer potencies for all carcinogenic chemicals at high doses is the same as the rank order of cancer potencies at all lower doses. As Peto *et al.* note,[20] given the assumed consistency of rank, the specification of a particular dose at which to calculate the index is arbitrary. By analogy with acute toxicity, they[18–20] propose the median response dose. They note the convenience of this choice in that most experimental dose ranges include this value. However, they also note that similar statistical arguments could be made about the TD_{10} or TD_{01}.

[4] Median response is a term we shall use to describe the dose at which half of those exposed exhibit a response after adjusting for background rates of response and intercurrent mortality. Quantal responses discussed later in the text refer to those doses at which the specific quantile of the exposed population is expected to exhibit a response.

4. DIFFERENT DOSE–RESPONSE FUNCTIONS

One limitation of using the TD_{50} to rank carcinogens is that there is an implicit assumption that data from all suspect chemicals be fit with the same dose–response model and that the TD_{50} be inferred from that model.[20] Others have shown that dose–response curves with different shapes (e.g., probit, logistic and one-hit) and identical TD_{50} and TD_{16} values have very different $TD_{0.0001}$ values.[11] With models of less similar functional forms, even greater variations would be expected at low exposures from chemicals with identical TD_{50} values. However, using a single functional form to model responses to all chemicals implies that cancer is a single process, or at least that macroscopically it works by a single set of rate kinetics. There is no allowance for biological differentiation or alternative pathways. We view this constraint as unlikely.

5. EQUALITY OF VARIANCES

The second assumption implicit in using the TD_{50} is that responses are modeled as having the same variance, irrespective of the value of the TD_{50}. Consider the probit-log dose model. Under this model, the probability of an individual's response is modeled as a normal probability distribution with respect to the log of the dose. The cumulative form of this function is fit to experimental animal data. In terms of the probability distribution, two parameters are fit to this functional form, the central tendency and the rate of spread from the center. Implicit in the TD_{50} approach is the assumption that chemicals with the same TD_{50}, the central tendency of this distribution, have the same rate of spread through the tails, the second moment of the distribution or variance (which is analogous to the slope of the cumulative distribution curve). That is, two curves with the same TD_{50} would be identical at all doses. In mathematical terms, this means that the ratio of TD_{50} to $TD_{0.0001}$ should be constant over all chemicals. Not only is this assumption biologically implausible, it is not supported by extant data.

In graphical terms, the curves for two chemicals with different TD_{50}s would be identical but equidistantly

displaced at all points. Figure 1 shows three hypothetical dose–response curves fit to the probit model. Frame A shows the distribution of individual susceptibilities. The distribution follows a normal probability distribution. Curves 1 and 2 have the same variance, or rate of change, but different TD_{50}s. Curves 2 and 3 have different variances but the same TD_{50}s. The cumulative versions of these curves are shown in Frame B. These are the data that typically are fit to animal bioassay data. Curves 1 and 2 in Fig. 1B are equidistant (along the x-axis for any value along the y-axis). Curve 3 has a different rate of spread from Curves 1 and 2 and thus intersects them. Since we are using a probit model, the data can be subjected to an inverse normal transform to yield straight lines, as in Fig. 1C. In this case, only parallel lines fit

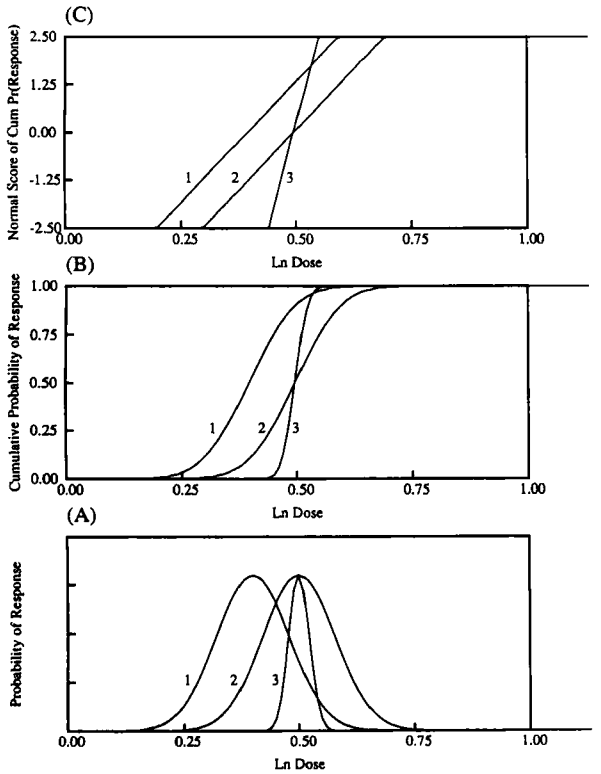

Fig. 1. (A) Three hypothetical dose–response curves for three different chemicals fit to probit models. This figure shows the distribution of susceptibilities in a population to different compounds as a lognormal distribution. Curves 1 and 2 have the same variance but a different TD_{50}. Curves 2 and 3 have different variance but the same TD_{50}. (B) The cumulative distributions of the curves in A. This is the form in which the information is presented most often. (C) Inverse normal (z-score) transform of the cumulative distributions in B. The intersection of Curve 3 with Curves 1 and 2 shows that the rank order of potencies changes as a function of dose. Note that Curves 1 and 2 have the same slope (variance), and Curves 2 and 3 have the same midpoint (TD_{50}).

the Ames *et al.* model (Curves 1 and 2). Nonparallel lines (Curve 3 with respect to 1 and 2) violate their implicit assumption. When these lines cross, their rank order changes, a condition not considered in the Ames *et al.* model. Rather than being a rare, statistical observation, this rank change is common in extant data. Similar problems occur if one considers the TD_{50} and other two or more parameter models of dose–response.

Consider, for example, a sample of data presented by these same authors.[23] We have chosen two compounds to demonstrate the problem: AF-2 (no. 99) and vinyl chloride (no. 2867). Comparison of the dose–response curves (Table I, Fig. 2) shows that the rank order of responses is not preserved from the TD_{50} to the range of values tested. While vinyl chloride has the larger TD_{50} (843 vs. 225), meaning that it is less toxic, AF-2 is less toxic at doses of about 100 mg/kg and below (1/50 vs. 13/60). Both differences, taken independently, are statistically significant at $p \leq 0.01$. For this one arbitrary pair of chemicals, the TD_{50} concept does not correctly order cancer potencies at low dose. Not only is the Ames *et al.* concept theoretically flawed, but it does not work for these empirical data. Other similar examples exist.

In general, the rank reversal is related both to the shape of the dose–response curve (the slope and the rate of change of the slope) and carcinogenic potency at high-dose exposures rather than the high-dose potency alone. In fact, substances with supralinear dose–response curves (flatter on a log–log plot) below the TD_{50} (e.g., vinyl chloride) will have greater ratios of the $TD_{50}/TD_{0.0001}$ than those with sublinear curves (steeper on a log–log plot—e.g., AF-2). We suggest that the differences in slopes and rates of change of slopes may be due to different biological processes involved in the activation and detoxification of these different materials.

Of additional importance is the reliability of the data. While, for our example, the 95% confidence intervals of the TD_{50}s do not overlap, one may question whether there is any significant difference in carcinogenic potency at any dose between these two substances. Statistically, they are different although this difference may not be meaningful or interesting biologically or in terms of hazard. Such considerations bring into question the overall reliability of QRA, which is not the subject of this note. However, we caution that the uncertainties in the animal bioassay results suggest that comparisons using any statistical measure of potency at any dose of these compounds is tenuous. One should exercise extreme caution in comparing compounds of similar potencies at any dose or with intersecting empirical dose–response curves. For compounds that have markedly different dose–response curves or different mechanisms of

Table I.

AF-2			Vinyl chloride		
Dose (mg)	*P(d)*	*P(d) − P(0)*	Dose (mg)	*P(d)*	*P(d) − P(0)*
0	0/65	0.00	0	0/58	0.00
104	1/50	0.02	2.02	1/60	0.02
520	25/50	0.50	10.10	3/59	0.05
			20.20	6/60	0.10
			101.00	13/60	0.22
			243.00	13/59	0.22
225		TD_{50}	843.00		TD_{50}

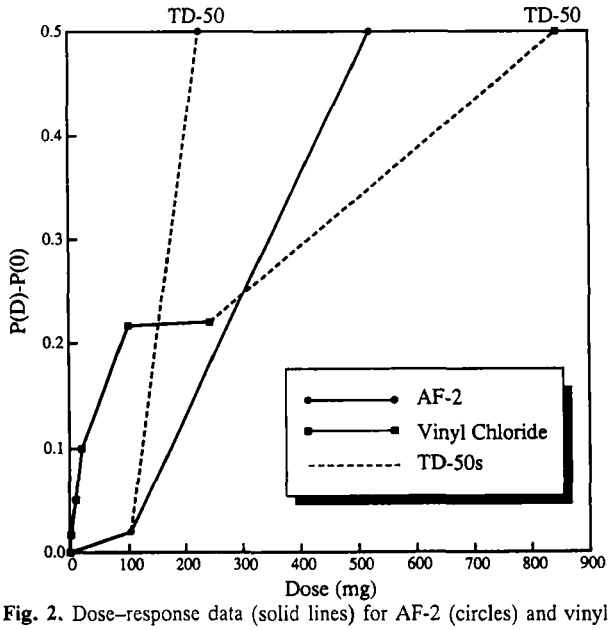

Fig. 2. Dose–response data (solid lines) for AF-2 (circles) and vinyl chloride (boxes) from Gold *et al.*[23] The TD_{50} (dashed lines) for vinyl chloride is greater than that for AF-2, suggesting that AF-2 is more carcinogenic at doses higher than those tested. However, at the lower doses tested, vinyl chloride is more carcinogenic that AF-2. Thus, for this one pair of chemicals, the TD_{50} concept does not correctly order cancer potencies at low dose. This same reversal is seen for other chemicals and for other dose–response models with the same chemicals (see text).

action, any measure of potency will document their difference, at least qualitatively.

6. DISCUSSION AND CONCLUSIONS

Ames *et al.*[18–20] have proposed the use of relative carcinogenic potency rather than QRA in assessing human hazard to environmental exposures. They suggest using the TD_{50}, which is an estimate of carcinogenic effect among test animals at high doses, to evaluate chronic human health hazard at low doses. While noting the limitations of their new index (e.g., differential susceptibility between rodents and humans, unknown dose–response functions, different mechanisms of action among carcinogens), they use it without modification. Yet, since dose–response curves may differ in shape, location, and spread, the relevance of their TD_{50} for predicting human response at low levels of environmental exposure is problematic.

Other recent evaluations of health using QRA have taken great strides toward incorporating some biological evaluation into the dose–response modeling.[14] At their most sophisticated, such efforts have included physiologically based pharmacokinetic models, evaluations that include biological measures of reaction and response within different systems of the test animal. We believe that this incorporation of biological information into the statistical evaluations is the appropriate direction for these models to grow and that the TD_{50} concept will hinder such advancement.

We have argued that unless the same functional form is used to model dose–response functions, and unless equality of variances is assumed for responses to different chemicals, it is unlikely that the rank order of potencies will be preserved over the wide range of exposures considered by Ames *et al.*[18–20] Further, the constraints proposed by them preclude the incorporation of biologically plausible components into dose–response evaluations. This argues strongly against the use of the TD_{50} (or any other single-parameter relative potency index) because the strict conditions for its reliability are statistically unlikely and because it excludes consideration of biological mechanisms.

Finally, Ames *et al.*[18–20] suggest the use of the TD_{50} for priority setting in the management of human

health risks. We suggest that adverse health outcomes other than cancer should be included in any evaluation of human hazard. For example, both cases of contaminated water that Ames *et al.* cite were of concern because of adverse reproductive outcome as well as cancer. Pesticides are known to affect the reproductive, neurological, and immunological systems as well as being carcinogenic. In general, risks of adverse reproductive outcome, of neurologic and immunologic dysfunction, and other impairments must be considered along with cancer in human health risk evaluation. We realize that the risk assessment methods for studying these problems are less well-developed than for cancer and that the database is more sparse. However, these other outcomes may be more sensitive indicators of human health hazard.

ACKNOWLEDGMENTS

We thank Marvin A. Schneiderman and two reviewers for helpful and insightful comments.

REFERENCES

1. W. R. Bryan and M. B. Shimkin, "Quantitative Analysis of Dose–Response Data Obtained with Three Carcinogenic Hydrocarbons in Strain C3H Mice," *Journal of the National Cancer Institute* **3**, 503–531 (1943).
2. N. Mantel and W. R. Bryan, " 'Safety' Testing of Carcinogenic Agents," *Journal of the National Cancer Institute* **27**, 455–470 (1961).
3. J. Cornfield, "Carcinogenic Risk Assessment," *Science* **198**, 693–699 (1977).
4. D. Krewski and J. Van Ryzin, "Dose–Response Models for Quantal Response Toxicity Data," in *Statistics and Related Topics*, M Csorgo, D. A. Dawson, J. N. K. Rao, A. K. Md. E. Saleh (eds.) (North-Holland, 1981), pp. 201–231.
5. P. Armitage and R. Doll, "A Two-Stage Theory of Carcinogenesis in Relation to the Age Distribution of Human Cancer," *British Journal of Cancer* **11**, 161–169 (1957).
6. J. Neyman and E. Scott, in *Fifth Berkeley Symposium on Mathematical Statistics and Probability* (University of California Press, Berkeley, 1967), pp. 745–776.
7. K. S. Crump, D. G. Hoel, C. H. Langley and R. Peto, "Fundamental Carcinogenic Processes and Their Implications for Low-Dose Risk Assessment," *Cancer Research* **26**, 2973–2979 (1976).
8. M. C. Pike, "A Method of Analysis of a Certain Class of Experiments in Carcinogenesis," *Biometrics* **22**, 142–161 (1966).
9. K. G. Brown and D. G. Hoel, "Modeling Time-to-Tumor Data: Analysis of the ED_{01} Study." *Fundamentals of Applied Toxicology* **3**, 458–469 (1983).
10. J. D. Kalbfleisch, D. Krewski, and J. Van Ryzin, "Dose-Response Models for Time to Response Toxicity Data (with discussion)." *Canadian J. Stat.* **11**, 25–49 (1983).
11. N. Chand and D. G. Hoel, "A Comparison of Models for Determining Safe Levels of Environmental Agents," in *Reliability and Biometry: Statistical Analysis of Lifelength* (SIAM, Philadelphia, 1974), pp. 681–700.
12. D. G. Hoel, "Incorporation of Background in Dose–Response Models," *Federation Proceedings* **39**, 73–75 (1980).
13. D. Krewski, K. S. Crump, J. Farmer, D. W. Gaylor, R. Howe, C. Portier, D. Salsburg, R. L. Sielken, and J. Van Ryzin. "A Comparison of Statistical Methods for Low Dose Extrapolation Utilizing Time-to-Tumor Data," *Fundamentals of Applied Toxicology* **3**, 140–160 (1983).
14. *Pharmacokinetics and Risk Assessment. Drinking Water and Health*, Vol. 8 (National Academy of Sciences, Washington, D.C. 1987).
15. S. H. Moolgavkar and D. J. Venzon, "Two Event Models for Carcinogenesis: Incidence Curves for Childhood and Adult Tumors," *Mathematical Biosciences* **47**, 55–77 (1979).
16. S. H. Moolgavkar and A. G. Knudson, Jr., "Mutation and Cancer: A Model for Human Carcinogenesis," *Journal of the National Cancer Institute* **66**, 1037–1052 (1981).
17. T. W. Thorslund, C. C. Brown, and G. Charnley, "Biologically Motivated Cancer Risk Models," *Risk Analysis* **7**, 109–119 (1987).
18. B. N. Ames, R. Magaw, and L. S. Gold, "Ranking Possible Carcinogenic Hazards," *Science* **236**, 271–280 (1987).
19. C. Sawyer, R. Peto, L. Bernstein, and M. C. Pike, "Calculation of Carcinogenic Potency from Long-Term Animal Carcinogenesis Experiments," *Biometrics* **40**, 27–40 (1984).
20. R. Peto, M. C. Pike, L. Bernstein, L. S. Gold, B. N. Ames, "The TD_{50}: A Proposed General Convention for the Numerical Description of the Carcinogenic Potency of Chemicals in Chronic-Exposure Animal Experiments," *Environmental Health Perspectives* **58**, 1–8 (1984).
21. J. Tierney, "Not to Worry," *Hippocrates* **29**, 29–38 (1988).
22. C. J. Portier and D. G. Hoel, "Issues Concerning the Estimation of the TD_{50}," *Risk Analysis* **7**, 437–447 (1987).
23. L. S. Gold, C. B. Sawyer, R. Magaw, G. M. Backman, M. de Veciana, R. Levinson, N. K. Hooper, W. R. Havender, L. Bernstein, R. Peto, M. C. Pike, and B. N. Ames, "A Carcinogenic Potency Database of the Standardized Results of Animal Bioassays," *Environmental Health Perspectives* **58**, 9–319 (1984).