Risk Analysis, Vol. 12, No. 1, 1992

Monte Carlo Simulation of Rodent Carcinogenicity Bioassays

Alexander Shlyakhter,[1,4] Gay Goodman,[1–3] and Richard Wilson[1]

Received December 18, 1990; revised May 17, 1991

In this paper we describe a simulation, by Monte Carlo methods, of the results of rodent carcinogenicity bioassays. Our aim is to study how the observed correlation between carcinogenic potency (β or $\ln 2/TD_{50}$) and maximum tolerated dose (MTD) arises, and whether the existence of this correlation leads to an artificial correlation between carcinogenic potencies in rats and mice. The validity of the bioassay results depends upon, among other things, certain biases in the experimental design of the bioassays. These include selection of chemicals for bioassay and details of the experimental protocol, including dose levels. We use as variables in our simulation the following factors: (1) dose group size, (2) number of dose groups, (3) tumor rate in the control (zero-dose) group, (4) distribution of the MTD values of the group of chemicals as specified by the mean and standard deviation, (5) the degree of correlation between β and the MTD, as given by the standard deviation of the random error term in the linear regression of log β on log (1/MTD), and (6) an upper limit on the number of animals with tumors. Monte Carlo simulation can show whether the information present in the existing rodent bioassay database is sufficient to reject the validity of the proposed interspecies correlations at a given level of stringency. We hope that such analysis will be useful for future bioassay design, and more importantly, for discussion of the whole NCI/NTP program.

KEY WORDS: Animal bioassays; Monte Carlo simulations; chemicals; carcinogens; correlations.

1. INTRODUCTION

The extrapolation of carcinogenic potencies from animals to man is rendered plausible if it can be shown experimentally that it is possible to predict a chemical's potency in one animal species given the potency in another species. To this end, Crouch and Wilson,[1] Crouch *et al.*,[2] and Crouch[3] analyzed the relationship between the carcinogenic potencies of chemicals in rats and mice, and Gold *et al.*[4–6] created the Carcinogenic Potency Database to facilitate future comparisons.

There have been several analyses of the correlation between carcinogenic potencies in rats and mice, as has been elaborated in a recent review by Goodman and Wilson.[7] But the basis and relevance of this correlation have been the subject of considerable discussion. To "predict" the likelihood that humans will develop cancer following a given level of exposure to specific chemicals, society inevitably relies upon indirect methods. The usefulness of rodent bioassay data for quantitative prediction of human carcinogenicity depends upon satisfactory treatment of two major uncertainties: (1) the interspecies comparison of carcinogenic potency when the dose levels are similar; and (2) the ability to predict low-dose tumorigenicity from high-dose experiments.

In this paper we examine the first point by means

[1] Harvard University, Department of Physics, Cambridge, Massachusetts, 02138.
[2] School of Public Health, 677 Huntington Avenue, Boston, Massachusetts, 02115.
[3] Present address: Gradient Corporation, 44 Brattle Street, Cambridge, Massachusetts 02138.
[4] To whom all correspondence should be addressed.

0272-4332/92/0300-0073$06.50/1

of a Monte Carlo simulation. We first make certain assumptions about the system under consideration, which in this case is a set of possibly carcinogenic chemicals. We then simulate the results of bioassays in the NTP program. Foremost among these assumptions is that the number of chemicals under test form a random sample. Taking these simulated results, we then perform linear regression analysis for β regressed on MTD (analogously to Crouch and Wilson[1] and other authors). We repeat this for a variety of initial assumptions and compare the simulated results of experiments to the actual results, so that we may discover which assumptions can be ruled out by the actual results of the experiments, and which are possible descriptors of the real world. In the performance of the animal bioassay, the toxicity of the chemical determines the upper limit on the dose level: the maximum dose level (i.e., the MTD) is adjusted so as not to produce overt toxicity or significant weight loss. As other authors have pointed out,[8,9] the fact that the upper limit on the dose is constrained by toxicity effectively imposes an upper limit on the carcinogenicity values. Therefore, it places an artificial constraint on the results which we must model in our simulation procedure.

We perform the calculation in two parts. First, we address the correlation between carcinogenic potency and the maximum tolerated dose and show how constraints imposed by the bioassay design (such as that the dose approach, but not exceed, the MTD) could with some parameters produce a correlation which is a pure artifact of these experimental constraints, but by comparison with the actual data show that the experimentally observed correlation cannot be an artifact of the constraints and therefore must have some biological or chemical basis.

Second, we perform a more complex simulation of two bioassays (e.g., one for rats and one for mice) and show that the observed correlation of interspecies potencies cannot be attributed only to the design of the bioassays and therefore be properly labeled artifactual. At least some fraction of this correlation is attributable to the biological similarity of the two species.

2. METHODS

2.1. Details of the Simulation Procedures

In this section we describe how the correlation between the carcinogenic potency (β) and the toxicity parameter A_i ($A_i = 1/\mathrm{MTD}_i$) is simulated. Throughout the paper, β and MTD are measured in units of mg/kg-day. We assume that there are a number of chemicals 1, ...i, ...n, and that A_i is a measure of toxicity of the i^{th} chemical. In this simulation, we assume a lognormal distribution of A_i with $x = \log_{10}A$, mean $\bar{x}$, and standard deviation σ_1.

$$P(x)=\frac{1}{\sqrt{2\pi}\sigma_2}e^{-\frac{(x-\bar{x})^2}{2\sigma_1^2}} \tag{1}$$

We then assume there is some relationship between the carcinogenic potency β_i of chemical i and its toxicity. If that relation was completely deterministic [$\beta = f(A)$], we could simply determine the distribution of the potencies β_i from the distribution of toxicities A_i.

The Monte Carlo approach is useful for considering cases where the relation is *not* completely deterministic [i.e., where β is only "roughly equal" to $f(A)$]. We use the formula

$$2.3\log_{10}\beta_i = 2.3b_1\log_{10}A_i + b_0 + \epsilon_i \tag{2}$$

where ϵ_i is a random error term. Note that we use the natural logarithm of A in this equation. We assume that ϵ is normally distributed:

$$P(\epsilon)=\frac{1}{\sqrt{2\pi}\sigma_2}e^{-\frac{\epsilon^2}{2\sigma_2^2}} \tag{3}$$

In this equation, when the adjustable parameter $\sigma_2 \approx 0$, $\epsilon \approx 0$, and the assumed relation between β_i and A_i is then nearly exact. When $\sigma_2 \geq 2$, β_i and A_i are virtually uncorrelated.

For $b_1 = 1$, the potency β_i is proportional to the toxicity parameter A_i (inversely proportional to MTD_i). If ϵ_i is small, this proportionality is nearly exact for each chemical. In the set of simulations in this paper, b_1 is chosen equal to 1.0.

Once β_i is determined from Eq. (2), the probability that an animal will get cancer at a given dose d is determined from the simple formula

$$P_i(d)=1-(1-a_0)e^{\frac{\beta_i d}{1-a_0}} \tag{4}$$

Here a_0 is the tumor rate in the zero-dose group. Although this formula is linear in dose at low doses, we emphasize that the results will not be appreciably changed by using more complex formulae which give similar results for P_i between 0.25 and 0.75 but give different results at low doses. Using Eq. (4) we can calculate the tumor probability for a single animal at any one of the dose levels $d = 0$, $\mathrm{MTD}_i/(N_d-1)$, $2\cdot\mathrm{MTD}_i/(N_d-1)$, ... MTD_i. We assume that the highest dose is MTD_i and N_d is the number of dose groups including the control group. For most chemicals in our dataset, there were

only two dose groups and one control group, so this is the number that we chose to use in most of the simulations. Then we use the binomial distribution to obtain the probability that exactly k animals will get tumors, and finally the cumulative distribution function $F_i(d)$, which gives the probability that the number of animals with tumors is less than or equal to k. Values of this function F_i lie in the interval (0,1). Using random sampling, we choose a point in this interval and derive the number of animals with tumors, $T_i(d)$, in the "simulated" experiment corresponding to this value of $F_i(d)$. This completes the first step.

The next step is to take this "simulated" number of cancers $T_i(d)$ and use it to calculate the experimental simulated potency, β_i^*, by the maximum likelihood estimate (MLE) method in the same way as was done by Crouch and Wilson.[1] This will be different from the originally assumed β_i because of randomness introduced into the simulated experiment. The value of β_i^* may or may not be statistically significant, depending on the criterion chosen for significance and the constants b_1 and b_0 in the original relation between β and A. Since data which are not statistically significant are generally ignored, this leads to a biased rejection of chemicals for which β_i is small relative to $1/\text{MTD}_i$. Here we assume that the chemical's potency is statistically significant if the maximum likelihood estimate β_i^* is at least two standard deviations away from zero (95.4% confidence level).

The final step in the simulation is to compare β_i^* with the initial A_i for each statistically significant chemical in order to see how the "measured" relationship between β and A compares with that initially assumed. The difference is the bias that we are trying to demonstrate and illustrate.

We repeat this procedure for each chemical i for all n chemicals in order to derive the full distribution in the finally measured β_i. As a measure of the strength of the relationship between $\ln(\beta)$ and $\ln(A)$, we compute r, the Pearson linear correlation coefficient, and s, the estimate of the standard deviation σ [Eq. (3)].

2.2. Dataset Selection

The dataset with which we compare our Monte Carlo simulations consists of a set of 248 chemicals which had been tested for carcinogenicity in rats and mice of both sexes by the NCI/NTP and included in the Carcinogenic Potency Data Base (CPDB) of Gold *et al.*[4–6] The TD_{50} values calculated by Gold *et al.* were translated into β values by assuming $\beta = \ln 2/\text{TD}_{50}$. For each chemical we selected the minimum TD_{50} which was statistically significant at $p < 0.05$ and assumed that the maximum dose (MaxD) used in that experiment was equal to the MTD. If the tumor incidence in the control group for a given site was greater than 60%, then the TD_{50} at that site was disregarded. Data from combined sites (reported as "tumor-bearing animals") were ignored. Only oral and inhalation routes were considered; therefore, we excluded 10 chemicals for which only intraperitoneal injection studies were available. Data for each sex in mice and rats were analyzed separately. The resulting dataset contained $248 - 10 = 238$ chemicals.

This dataset had been assembled initially by one of us (GG) for a study of the effect of mutagenicity on the correlation between toxicity and carcinogenicity.[10] No other selection criteria were used. In most cases there were two dose groups plus a control group with 50 animals in each. We also took note of the existence of any substances that produced tumors in 98% or more of the tested animals for analysis of the effect of the artificial constraints. There is only one such substance: polybrominated biphenyl mixture (Firemaster FF-1 CAS no. 67774-32-7). It produces tumors in 98% of the animals at the highest dose, but not at the intermediate dose. A full listing of the dataset, the number of chemicals tested, the number for which there existed a TD_{50} value significant at $p < 0.05$ for each sex of both species, and fraction of chemicals in each sex-species subgroup for which there was no statistically significant TD_{50} value in the CPDB are available upon request from the first author.

We chose the data for female mice as the basic set for comparison with the Monte Carlo simulations. Similar results were obtained when we used the slightly smaller dataset of Crouch.[3] Ideally, we should have analyzed the raw data on tumors for carcinogenic potency β by the same MLE as performed in the simulation. However, it was convenient to use our computer file of TD_{50} values calculated by Gold *et al.*

3. RESULTS AND DISCUSSION

First, we illustrate how the simulation procedure works. Then we study the sensitivity of the results to the parameters and choose the basic parameter set. Using this basic set we study the effect of allowing for the existence of a variable number of chemicals producing tumors in more than 98% of test animals, and show how Monte Carlo simulation can be used for optimization of the bioassay design. Finally, we simulate the interspecies correlation of carcinogenic potencies and the case of a variable number of chemicals having zero carcinogenic potency.

3.1. Illustration of the Simulation Procedure

In Fig. 1, we show a histogram of $\log_{10} A$, measured in mice, for the 238 chemicals in our dataset. The distribution of $\log_{10} A$ calculated with Eq. (1) is given as the smooth curve in Fig. 1. Figure 2 illustrates the properties of the artificial dataset that we later use as the input for the Monte Carlo simulation. For this dataset, 238 artificial "chemicals" were generated using the model for the dependence of carcinogenic potency β on $\log_{10} A$ given in Eq. (2) with slope $b_1 = 1.0$ and intercept $b_0 = -2.0$. This we call simulation no. 0, which is not really a simulated bioassay at all, but rather the input dataset for the simulated bioassays to follow. For each "chemical," values of $\log_{10} A$ were generated using the probability density function of Eq. (1) with the standard deviation σ_1 set equal to 1.0. Then β was calculated using this value of $\log_{10} A$ with a randomly generated value of the error term ϵ. Values of ϵ were generated using the probability density function of Eq. (3) with the standard deviation σ_2 set equal to 2.0. Each chemical is represented by a cross in Fig. 2. The solid line represents the linear least-squares regression fit to the data points.

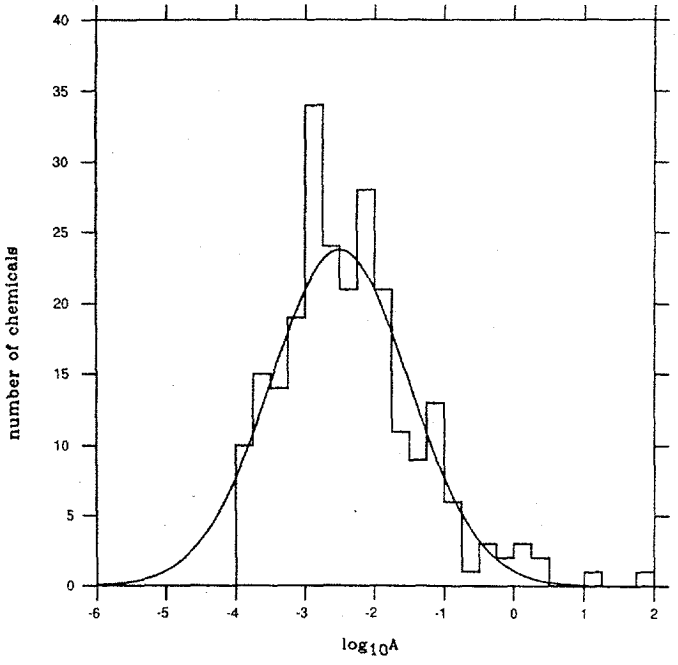

Fig. 1. *Distribution of $\log_{10}$ A measured in mice for 238 NCI/NTP chemicals.* The histogram represents the experimental distribution of $\log_{10} A \equiv \log_{10}(1/\text{MTD})$. MTD is measured in units of mg/kg−day. The mean $\bar{x}$ of this distribution was −2.36; the standard deviation σ_1 was 0.94. The solid line is the normal curve of Eq. (1) representing a continuous distribution of $x = \log_{10} A$ with the rounded parameters $\bar{x} = -2.5$ and $\sigma_1 = 1.0$.

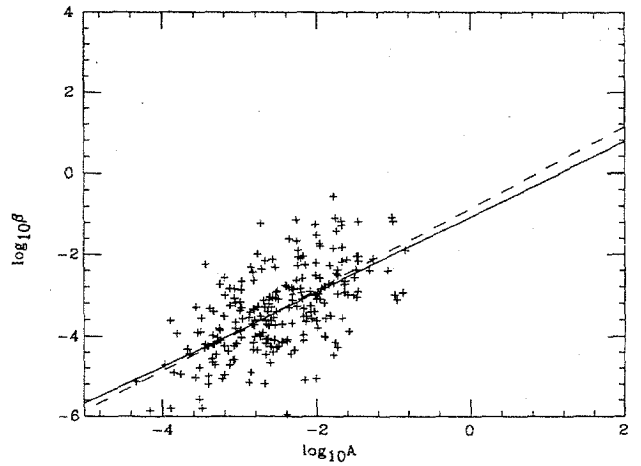

Fig. 2. *Artificial distribution with β calculated from Eq. (2).* The datapoints (+) were generated using Eq. (2) with randomly generated values of $\log_{10} A$ and the error term ϵ (see text for details). Solid line: linear regression fit to the datapoints. Dashed line: plot of Eq. (2) for $\epsilon = 0$.

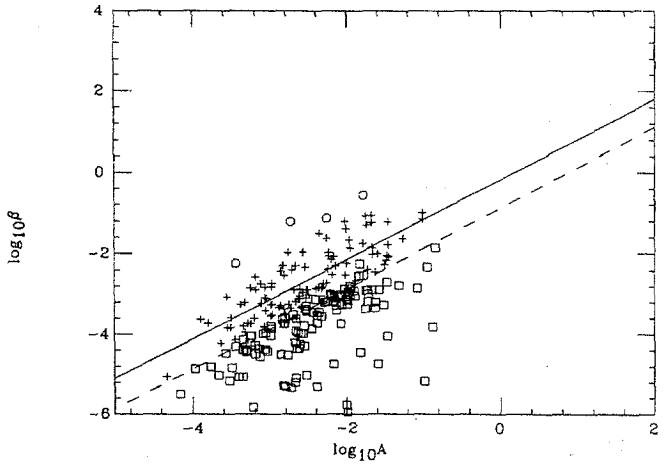

Fig. 3. *Results of a simulated bioassay (simulation no. 1).* +, β^* significant at $p<0.05$; percentage of animals with tumors is <98%. ○, $\beta^* = \beta$; precentage of animals with tumors is <98%. □, β^* not significant at $p<0.05$. Solid line: linear regression fit to the crosses. Dashed line: linear regression fit corresponding to the artificial dataset of Fig. 2.

To illustrate the effect of the error term ϵ, we also plot Eq. (2) for the case when the error term $\sigma_2 = 0$ (dashed line).

Figure 3 illustrates how the Monte Carlo simulation procedure works. Shown are the results of a simulated bioassay of the 238 artificial "chemicals" described above for Fig. 2. This we call simulation no. 1 in Table I. For each "chemical," the simulated experiment uses the

Table. I. Sensitivity to the Parameters[a]

	Input							Output			
Sim. no.	N_a	N_d	a_0	$\bar{x}$	σ_1	σ_2	b_0	NS	>98%	r	s
0	—	—	—	—	1.0	2.0	−2.0	122	1	0.87	0.59
1	50	3	0.1	−2.5	1.0	2.0	−2.0	128	4	0.78	0.54
2	25	3	0.1	−2.5	1.0	2.0	−2.0	160	6	0.82	0.48
3	100	3	0.1	−2.5	1.0	2.0	−2.0	99	4	0.75	0.56
4	200	3	0.1	−2.5	1.0	2.0	−2.0	79	4	0.73	0.59
5	50	2	0.1	−2.5	1.0	2.0	−2.0	132	12	0.79	0.52
6	50	4	0.1	−2.5	1.0	2.0	−2.0	119	4	0.80	0.53
7	50	3	0.1	−1.5	1.0	2.0	−2.0	128	4	0.78	0.54
8	50	3	0.1	−2.5	0.5	2.0	−2.0	128	4	0.52	0.54
9	50	3	0.05	−2.5	1.0	2.0	−2.0	123	4	0.73	0.57
10	50	3	0.2	−2.5	1.0	2.0	−2.0	136	5	0.78	0.50
11	50	3	0.1	−2.5	1.0	0.07	−2.0	148	0	0.99	0.11
12	50	3	0.1	−2.5	1.0	3.2	−2.0	129	36	0.60	0.96
13	50	3	0.1	−2.5	1.0	2.0	−1.0	75	12	0.77	0.55
14	50	3	0.1	−2.5	1.0	2.0	−3.0	166	1	0.85	0.45
15	50	3	0.1	−2.5	1.0	2.0	−2.0	128	4	0.82[b]	0.47[b]
16	50	3	0.1	−2.5	1.0	6.4	−2.0	123	53	0.44	1.75
17	50	3	0.1	−2.5	1.0	6.4	−2.0	123	53	0.84[b]	0.48[b]
18[c]	25	6	0.1	−2.5	1.0	2.0	−2.0	126.4 ± 3.2[d]			
19[c]	50	3	0.1	−2.5	1.0	2.0	−2.0	104.4 ± 2.3[d]			
20[c]	75	2	0.1	−2.5	1.0	2.0	−2.0	93.4 ± 8.6[d]			

[a] Unless otherwise stated we use $N_c = 238$ and study the sensitivity to the parameters, holding $b_1 = 1.0$. N_c, number of chemicals in simulation; N_a, number of animals per dose group; N_d, number of dose groups (including zero dose group); a_0, fraction of animals with background tumors; $\bar{x}$, mean value of $\log_{10} A$; σ_1, standard deviation of $\log_{10} A$ [Eq. (2)]; σ_2, standard deviation of the error term ϵ in [Eq. (3)]; b_0, parameter in Eq. (2); NS, number of "nonsignificant" chemicals; >98%, number of chemicals with more than 98% tumors; r, correlation coefficient of regression fit; s, standard deviation of regression fit.

[b] Chemicals which produce tumors in more than 98% of test animals were excluded from this regression.

[c] $N_c = 200$

[d] NS was averaged over five separate simulations.

probabilities from Eq. (4) to compute the number of animals that have developed tumors and, from this number, the "experimental" simulated potency β^*, by the MLE method. The 114 "chemicals" that have statistically significant ($p < 0.05$) β values *and* produce tumors in ≤98% of the animals are shown by crosses. As can be seen simply by overlaying Fig. 3 on Fig. 2, they are slightly shifted compared to their original positions, reflecting the limited statistical power of the simulated bioassay in determining the potencies. Those "chemicals" which are predicted to produce tumors in more than 98% of the animals are shown as circles. These are not shifted in position because, unable to calculate β_i^* for these cases, we used the original values of β_i. The chemicals for which the simulated β^* values are statistically nonsignificant are shown as squares. Because of the greater magnitude of associated uncertainty, they are shifted even more from their original positions in Fig. 2 than are the chemicals for which β^* is significant.

The regression (solid) line which was the "output" of Fig. 2 is drawn as the dashed line in Fig. 3 and is the "input" for the experimental simulation. The solid line in Fig. 3 is the linear least-squares regression fit to the crosses (i.e., the chemicals with significant β^* values that produce tumors in ≤98% of the animals). Since the 128 chemicals which give nonsignificant values of β lie *below* the others, the regression (solid) line lies higher than the dashed line.

The parameters for the other simulations are also shown in Table I. Also listed are the correlation coefficient (r) and standard deviation (s) for each of the solid lines. As anticipated, the biased selection of chemicals in simulation no. 1 (Fig. 3) results in a higher correlation coefficient ($r = 0.78$ for regression 1 instead of 0.59

for regression 0) and less deviation from the regression line ($s = 0.54$ instead of 0.87) than is obtained from the unconstrained data set of simulation no. 0 (Fig. 2).

3.2. Sensitivity to the Parameters

In order to choose which parameter sets give possible fits to the data, we studied how the results depend upon parameters. Since we assume the same parameter values in Eqs. (2) and (3) for each chemical and choose the doses for each chemical i proportional to its MTD_i, it follows from Eq. (5) that our results do not depend on the mean value of $\log_{10} A_i$. If the mean value of $\log_{10} A_i$ is changed, but nothing else, only the distribution as a whole is shifted. [Since $\beta_i = A_i \cdot \exp(b_0 + \epsilon_i)$ and $d = c \cdot \mathrm{MTD}$, where c is constant, then $\beta_i \cdot d = c \cdot \exp(b_0 + \epsilon_i) \cdot (A_i \cdot \mathrm{MTD})$ and does not depend on the values of A_i]. Variation of the standard deviation of $\log_{10} A$ (σ_1) (simulation no. 8) shows, as expected, an effect only on the linear correlation coefficient r, but not on the standard deviation s of $\ln(\beta)$ or the number of chemicals considered statistically nonsignificant.

One can see that when b_0 is varied from -1 to -3 [$\exp(b_0)$ varied from 0.368 to 0.050], shown in simulations no. 13 and 14, the carcinogenic potency is derived from Eq. (2) to be steadily smaller with respect to the MTD_i and the number of chemicals where potency is statistically nonsignificant increases from 75 – 166; but the number of chemicals producing tumors in more than 98% of the animals decreases from 12 to 1.

Figures similar to Fig. 3 were drawn for each of the simulations; they are available upon request.

3.3. Comparison of Simulated Experiments with the Real Experiment

We now take a final step; we compare the various artificial datasets with the experimental dataset to see whether the initial assumptions could possibly be correct.

With only two variable parameters, b_0 and σ, one cannot fit four parameters: NS, >98%, r, and s. We chose to fit NS, >98%, and s as closely as possible, since the first two are sensitive indicators of the potency vs. toxicity distribution and s gives the "band width." We now compare the parameters for the experimental dataset (close to simulation no. 0) and the simulated artificial datasets (simulations nos. 1–14 in Table II). A reasonable match is achieved with simulation no. 1 ($b_0 = -2.00$, $\sigma = 2.0$), for which NS = 128, >98% = 4, $r = 0.78$, and $s = 0.54$.

Table II. Sensitivity of Interspecies Potency correlation to Parameter σ_A[a]

	Input		Output			
Sim. no.	σ_A	NS	r^α	s^α	r^β	s^β
0[b]	—	61	0.94	0.38	0.84	0.67
21	0.07	58	1.00	0.06	0.70	0.64
22	0.21	58	0.97	0.19	0.69	0.65
23	0.35	58	0.91	0.32	0.67	0.70
24	0.53	58	0.83	0.48	0.63	0.77
25	0.71	58	0.74	0.64	0.59	0.87
26	1.41	58	0.48	1.28	0.42	1.39

[a] σ_A, standard deviation of the error term in Eq. (6); NS, number of "nonsignificant" chemicals; r^α, correlation of interspecies toxicities (from regression fit); s^α, standard deviation of interspecies toxicities (from regression fit); r^β, correlation of interspecies potencies (from regression fit); s^β, standard deviation of interspecies potencies (from regression fit).
[b] Correlations for NCI/NTP dataset.

Small variations of the parameters b_0 and σ_2 around $b_0 = -1.75$ and $\sigma_2 = 2.0$ (not shown here) do not improve the simultaneous fit to NS, >98%, s, and r. For example, realistic values of s are correlated with unrealistic values of >98% and vice versa. This is an indication that our model needs further refinement. At this stage we chose a compromise and use the set of parameters of simulation no. 1 in Table II as a basic one. For future study, one possible alternative to Eq. (4) could be to multiply the right-hand side by 0.9, describing a possibility that only 90% of animals ever get cancers. This would introduce a logical upper constraint on the data.

As noted before, the constraint of excluding cases with 98% or more of the animals developing tumors, used by Bernstein *et al.*,[8] is artificial and unnecessary. If any such case occurred in practice, no scientist would ignore it, and certainly no regulatory agency. We simulated the effect of including or not including this constraint, expecting to find a better correlation with it, which would indicate that it introduces a spurious correlation. For the basic parameter set (simulation no. 1), including the constraint results in only a small effect on the correlation coefficient r or standard deviation s (r rises from 0.78 – 0.82 and s falls from 0.54 – 0.47, simulation no. 15). This is not surprising, because in the actual dataset less than 2% of the chemicals yield more than 98% animals with tumors.

However, an artifactual correlation can arise when σ_2 is large enough. This is illustrated in simulation nos.

16, 17, where the basic set of parameters of simulation no. 1 is used, except for σ_2, which is made very large. In this case there are 53 chemicals which produce tumors in more than 98% of the animals at the MTD. Deletion of these chemicals decreases the number of chemicals with statistically significant potencies from 115 to 62. There is a good correlation if chemicals which produce tumors in more than 98% of the animals are deleted, but not if they are kept. The good correlation with these chemicals deleted is, therefore, clearly spurious and artifactual. It is interesting to compare the simulations no. 15, which uses the basic parameter set, and no. 17, which differs from 15 only in that it has a large σ. The values of r and s in the simulated data are very close in both simulations. This was the point stressed by Bernstein *et al.*,[8] and it shows that by looking at the constrained data only, one can tell little about the reality of the correlation. The shuffling method of reference 2 showed that the data in the bioassays is inconsistent with the proposition that there is no real correlation between carcinogenic potency in rats and mice. However, it did not claim to, and could not *by its nature*, show how constraints make measured correlation parameters differ from the true ones. That is the reason for this paper. However, we show here what Bernstein *et al.* did not: that the correlation is already contained in the following statements: (1) for nearly all chemicals, less than 98% of the animals get tumors; (2) about half of the chemicals have statistically nonsignificant potencies. The simulation of Bernstein *et al.* was less realistic than the one here: they assume that β·MTD is uniformly distributed between two limits (nonsignificance and 98% of animals with tumors), a sharp cutoff and no chemicals producing tumors in more than 98% of the animals. Our simulation uses a smoother distribution, and we perform the simulation before applying the constraints, not after applying them.

4. OPTIMIZATION OF BIOASSAYS

In simulation nos. 18, 19, and 20, only 200 chemicals were used, and the number of dose groups is varied from 6 to 2 while the number of animals per dose group is varied from 25 to 75 to keep the total number of animals constant at 150. The results were averaged over five independent simulations to provide an estimate of the uncertainty. The mean values and errors were estimated using the NS values obtained for each simulation.

Simulation no. 18 suggests that increasing the number of dose groups beyond a reasonable minimum value (while keeping the total number of animals constant) increases the number of nonsignificant chemicals appreciably; and simulation no. 20 shows that increasing the number of animals per dose group decreases that number. However, this conclusion depends critically upon the assumption of the dose–response relationship.

5. INTERSPECIES CORRELATION OF CARCINOGENIC POTENCIES

We finally examined the correlation between the carcinogenic potency in female rats ($\beta^{(r)}$) and the carcinogenic potency in female mice ($\beta^{(m)}$). The correlation between carcinogenic potencies in two different species is closely connected with the correlation between toxicity A and carcinogenic potency β in each of the species, and the existence of a correlation between toxicity in rats and mice ($A^{(r)}$ *vs.* $A^{(m)}$). As shown by Crouch *et al.*,[2] if three of the correlations exist, the fourth must exist also. If three of the correlations are exact, the fourth must also be exact. In this section we address the question whether the interspecies potency correlation that is observed is solely a consequence of the other three correlations or not.

We have performed some simulations of two simultaneous bioassays which may represent, for example, experiments on two different species: mice and rats. This is accomplished by assuming that MTDs (or MaxDs) in mice $A_i^{(m)}$ and in rats $A_i^{(r)}$ are related by equations similar to Eqs. (1) and (3):

$$\log_{10} A_i^{(r)} = \log_{10} A_i^{(m)} + \epsilon_{iA} \tag{5}$$

$$P(\epsilon_{iA}) = \frac{1}{\sqrt{2\pi}\sigma_A} e^{-\frac{\epsilon_{iA}^2}{2\sigma_A^2}} \tag{6}$$

The error term ϵ_A is normally distributed, with variance σ_A chosen to fit the data (see the dataset section above). Given a value of $\log_{10} A_i$, Eq. (2) is used to calculate a value of $\log_{10}\beta_i$. Estimated correlation coefficients and standard deviations are calculated for both $(A)^{(r)}$ *vs.* $(A)^{(m)}$ and $(\beta)^{(r)}$ *vs.* $(\beta)^{(m)}$. If we manage to fit the $(A)^{(r)}$ *vs.* $(A)^{(m)}$ distribution but the observed correlation of potencies appears stronger than the simulated one, we should conclude that the interspecies potency correlation contains additional information and is not just a consequence of the relation between $A^{(r)}$ and $\beta^{(r)}$ and between $A^{(m)}$ and $\beta^{(m)}$. It would not be a *statistical* artifact of the experimental design in any case since we have demonstrated that the correlations between A and β are not *statistical* artifacts of testing.

It is also interesting to analyze the interspecies cor-

relation of the number of "positive" (statistically significant) and "negative" (statistically nonsignificant) chemicals. If the interspecies potency correlation has some basis in biology and is not just a consequence of both the potency/toxicity correlation for each species and the interspecies toxicity correlation, one would expect that the observed number of chemicals that are "positive" (*i.e.*, their carcinogenic potencies are statistically significant at a given confidence level) in both species is higher than the number simulated.

We use our basic set of parameters (simulation no. 1) for simulating both the female mice and female rats datasets. Note that the additional error term ϵ_A increases the width of the lognormal distribution of toxicities in the second species. As we discussed earlier, this has no effect on the number of statistically nonsignificant chemicals or the parameter s; only the correlation coefficient r will be increased. However, r is already high and not very sensitive to the value of σ.

The results are presented in Table II. Here r^A and s^A refer to $(A)^{(r)}$ *vs.* $(A)^{(m)}$, and $r^{(\beta)}$and $s^{(\beta)}$ refer to $(\beta)^{(r)}$ vs. $(\beta)^{(r)}$ vs. $(\beta)^{(m)}$, respectively.

The value $\sigma_A = 0.35$ fits the observed correlation parameters (r^A and s^A) reasonably well. The interspecies potency correlation is stronger than any of the other three discussed in this paper, as is well known. The simulated interspecies correlation of carcinogenic potencies is somewhat weaker than the observed interspecies correlation of potencies. It is not yet clear whether this discrepancy proves that the observed interspecies correlation of potencies is enhanced relative to our simulation as a consequences of the toxicity/carcinogenicity relations, whether it is a statistical fluctuation, or whether it reveals some flaws of our model. Further work is needed to clarify the situation. Crouch *et al.*[2] pointed out that the interspecies correlation of carcinogenic potencies appears to be stronger than the correlation of toxicity and carcinogenic potency, and that this may have some (unknown) biological meaning.

We note that in these simulations we have assumed that the correlations $A^{(m)}$ *vs.* $\beta^{(m)}$ and $A^{(r)}$ *vs.* $\beta^{(r)}$ are statistically completely independent. One way of describing the fact that $\beta^{(r)}$ and $\beta^{(m)}$ are more closely correlated than either $A^{(m)}$ and $\beta^{(m)}$ or $A^{(r)}$ and $\beta^{(r)}$ is to postulate that if a chemical is more carcinogenic in rats than the correlation between A and β suggests, it is likewise more carcinogenic in mice than this correlation suggests. If the relationship between A_i and β_i is identical for rats and mice for each chemical i, then we expect the error term ϵ_{iA} to be identical with the error term $\epsilon_{i\beta}$ in an equivalent expression relating $\beta^{(r)}$ and $\beta^{(m)}$. It is clearly not, suggesting that the true situation lies between these two extremes.

6. HOW MANY NONCARCINOGENS CAN EXIST?

In the above simulations we assumed a loglinear relationship between carcinogenic potency β_i and toxicity A_i [Eqs. (2) and (3)] and in no case took $\beta \equiv 0$. This is in accord with our belief that it is wise to assume that all chemicals have some finite carcinogenic potency and that there is a "smooth" distribution of potencies, correlated in some way with toxicity. We note that this assumption is implicit in most of the published concordance and correlation studies.

Many toxicologists take another view, that there exist chemicals that are truly noncarcinogenic ($\beta_i = 0$) at any dose level. These we term "true" noncarcinogens. In the intermediate case, all chemicals would have nonzero carcinogenic potency if administered at high enough doses. Ames and Gold[11] recently reviewed the evidence that chemicals which produce cell proliferation due to cytotoxicity at high doses may be carcinogenic due to this mechanism only. These we term "threshold" noncarcinogens. It is easy to see how to simulate both the "true" and the "threshold" noncarcinogens. For example, we might assume that 41% (100) of all the 238 chemicals tested in female mice in the experimental dataset are true noncarcinogens ($\beta_i = 0$) and 9% (23) are carcinogens that have statistically nonsignificant potencies in female mice. This gives a total of 123 chemicals with either zero potency or statistically nonsignificant potency, and the remaining 115 have measurable, statistically significant, potency. For the "threshold" noncarcinogens, we assume that $\beta_i = 0$ only at MTD/2, but not at the MTD.

We then match these assumptions to the experimental datasets by performing two simulations, one (simulation no. 27) for which 41% of the 238 chemicals "threshold noncarcinogens" and one (simulation no. 28) for which 41% of the chemicals are "true noncarcinogens"; while only 9% of the chemicals are assumed to be carcinogens with statistically nonsignificant potencies. Simulations nos. 27 and 28 are shown along with the experimental dataset, simulation no. 0, in Table III.

We see that our parameter set is consistent with the existence of 41% "threshold" noncarcinogens (simulation no 27). However, assuming 41% "true" carcinogens (simulation no 28), we get too high a fraction of statistically nonsignificant chemicals. We then repeat the parameter fitting and, if a reasonable fit is not achieved

Table III. Effect of Noncarcinogens[a]

Sim.	Input							Output			
no.	N_a	N_d	a_0	$\bar{x}$	σ_1	σ_2	b_0	NS	>98%	r	s
0	50	3	0.1	−2.5	1.0	2.0	−2.00	122	1	0.87	0.59
27[b]	50	3	0.1	−2.5	1.0	2.0	−2.00	122	1	0.83	0.45
28[c]	50	3	0.1	−2.5	1.0	2.0	−2.00	174	1	0.75	0.53
29[c]	50	3	0.1	−2.5	1.0	1.0	−0.50	115	1	0.84	0.40

[a] Notation from Table I.

[b] 41% (100) chemicals assumed "secondary" noncarcinogens (*i.e.*, with nonzero potency at the MTD, and zero potency at MTD/2).

[c] 41% (100) chemicals assumed "true" noncarcinogens with zero potency.

for any choice of parameters, it indicates that such a fraction of noncarcinogens is incompatible with the data and can be excluded. One attempt is shown in Table III (simulation no. 29). It gives reasonable fits to the parameters NS, >98%, and r, but too low a value of s.

We ask if it is consistent with the observed interspecies potency correlation to assume that "true" noncarcinogens exist. The results of these simulations suggest that true noncarcinogens may exist but only if those chemicals that are truly carcinogenic have a closer correlation between toxicity (A) and carcinogenic potency (β) than what is experimentally observed for the 238 chemicals in our dataset.

7. SUMMARY

1. Algorithms and comprehensive software for Monte Carlo simulation of animal bioassays have been developed.
2. Simulations of carcinogenic potency *vs.* 1/MTD and simulations of the interspecies correlation suggest, as it was anticipated, that the initial correlation can be enhanced, or spurious correlations can arise, when constraints (such as limiting the percentage of animals with tumors to <98%) are added. However, for realistic values of the parameters (*i.e.,* those that fit actual data), there is already a strong correlation that arises due to the interspecies toxicity relation and the toxicity/potency relation; the observed correlation is only slightly stronger than what can be attributed to these two relations. Therefore, the effect seems to be insignificant. Important factors contributing to this conclusion are: (1) that there are very few chemicals for which more than 98% of the animals get tumors; and (2) for about half of the chemicals only a statistically nonsignificant number of animals get tumors. These two statements contain almost all our present information about the correlation.
3. Monte Carlo simulations can be helpful for planning future bioassays. For example, given a fixed number of animals for use in a bioassay, it is instructive to examine the influence of the number of dose groups and the number of animals per dose group on the number of chemicals that are found to have statistically significant carcinogenic potencies.
4. Monte Carlo simulations may turn out to be helpful in addressing such controversial issues as the interspecies potency correlation and the possible existence of noncarcinogens. However, further refinement and verification of our model are necessary before definite biological inferences can be drawn.

Since there has been a lot of confusion about the correlation between carcinogenic potency and toxicity, it may be useful to describe what we have *not* demonstrated.

We have assumed that the 238 chemicals in our dataset are a fair, random sample of chemicals of interest. While we have made no deliberate selection, we suspect that this may be the weakest of our assumptions.

We have barely discussed dose–response relationshiops. Our comparison of Monte Carlo calculations with the measured correlation have been with potencies mostly measured in the region where 10–90% of animals get tumors. Equation (4) implies a linear dose–response at low doses, which is almost certainly not true for some chemicals. Although our approach can be extended to nonlinear dose–response curves, we have not done so in this paper.

Our comparison of the Monte Carlo calculations with measured correlations only applies to the set of chemicals in teh NCI/NTP dataset. We do *not* address the representativeness of that set of chemicals.

ACKNOWLEDGMENTS

We appreciate the helpful comments on a draft of this paper by Myron Fiering, David Gaylor, Joseph Haseman, and Ronald Hart, and the assistance of Peter Cho and Robert McDonald with programming. We gratefully acknowledge gifts to Harvard University from Stone & Webster, Monsanto, Rohm & Haas, Gillette,

and Bristol-Meyers that supported this work. Richard Wilson was partially supported by a governmental interagency exchange agreement with FDA through NCTR.

REFERENCES

1. E. Crouch and R. Wilson, "Interspecies Comparison of Carcinogenic Potency," *J. Toxicol. Environ. Health* **5**, 1095–1118 (1979).
2. E. Crouch, R. Wilson, and L. Zeise, "Tautology or Not Tautology?" *J. Toxicol. Environ. Health 20*, 1–10 (1987).
3. E. A. C. Crouch, "Uncertainties in Interspecies Extrapolations of Carcinogenicity," *Environ. Health. Perspect.* **5**, 321–327 (1983).
4. 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," *Environ. Health Perspect.* **58**, 9–319 (1984).
5. L.S. Gold, T. H. Slone, G. M. Backman, R. Magaw, M. Da Costa, P. Lopipero, M. Blumenthal, and B. N. Ames, "Second Chronological Suplement 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).
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," *Envrion. Health Perspect.* **67**, 161–200 (1986).
7. G. Goodman and R. Wilson, "Predicting the Carcinogenicity of Chemicals in Humans from Rodent Bioassay Data," *Envrion. Health Perspect.* **94**, 195–218 (1991).
8. L. Bernstein, L. S. Gold, B. N. Ames, M. C. Pike, and D. G. Hoel, "Some Tautological Aspects of the Comparison of Carcinogenic Potency in Rats and Mice," *Fundam. Appl. Toxicol.* **5**, 79–86 (1985).
9. J. P. Rieth and T. B. Starr, "Experimental Design Constraints on Carcinogenic Potency Estimates," *J. Toxicol. Environ. Health* **27**, 287–296 (1989).
10. G. Goodman, A. Shlyakhter, and R. Wilson, "The Relationship Between Carcinogenic Potency and Maximum Tolerated Dose Is Similar for Mutagens and Non-mutagens," in B. E. Butterworth, T. J. Slaga, W. Farland, and M. McClain (eds.), *Chemically Induced Cell Proliferation: Implications for Risk Assessment* (Wiley-Liss, New York, 1991), pp. 501–516.
11. B. N. Ames and L. S. Gold, "Too Many Rodent Carcinogens: Mitogenesis Increases Mutogenesis," *Science* **249**, 970–971 (1990).