

Uncertainties in Interspecies Extrapolations of Carcinogenicity

by E. A. C. Crouch*

The usual classification of results of animal carcinogenicity tests is positive or negative. Attempting to observe correlations between species using such results is complicated by differing test sensitivities. In these circumstances it is helpful to use models to represent the experimental data in a consistent way. Fitting the model parameters to the data allows computation of confidence limits and an assessment of concordance or discordance between different species in a way which accounts for differing test sensitivities. This paper describes this approach in detail for one class of models applied to the carcinogenicity test results of 187 of the NCI bioassay series, allowing comparison between B6C3F1 mice, Fischer 344 rats and Osborne Mendel rats. It is shown that the uncertainties in extrapolating between species are larger than generally acknowledged (a standard deviation of a factor of ~ 4.5), but that within these uncertainties there are few if any discordances.

Introduction

When a test for carcinogenicity is performed on groups of experimental animals, the usual way of reporting the result is qualitative—as a positive when a dosed group of animals has a sufficiently large excess tumor rate or a negative if any excess tumor rates are not sufficiently large. The criterion for sufficiency is usually statistical, while the negativity of any result is usually qualified by the experimental conditions, among which are the dose of material under test and the numbers of animals tested. Attempts have been made to observe correlations between species by using such qualitative results (1), the first step in verifying efforts at extrapolation between species, but they are complicated by the differing experimental conditions and hence differing sensitivities. Furthermore, the measures of correlation which may be obtained from qualitative results are limited, and the application of these measures to interspecies extrapolation is unhelpful, since such extrapolations are usually required to be quantitative.

Under these circumstances, the use of models is helpful in condensing the considerable experimental data into a few convenient, quantitative parameters which may then be used as surrogates for all the experimental data. The definitions of the model and its parameters may be chosen to incorporate such

biases as the existence of a background rate for tumors, the expectation that tumor incidence rates rise as dose of carcinogen is increased, and the expectation that tumor incidence is stochastic. If the model is then applied to a large set of experiments on different materials in various species, the values of the model parameters obtained in each case represent uniform measures of the experimental results in each case, and are thus ideal candidates for observing correlations and for use in extrapolation. Moreover, it is usually possible to obtain some sort of measure of uncertainty and significance for such parameters, allowing a better assessment of concordance or discordance between experimental results.

The purpose of this paper is to give some details of this approach by using a particular model and to show the results obtained when this model is applied to the data generated by the National Cancer Institute (NCI) Carcinogenesis Bioassay program.

Method

The NCI Carcinogenesis Bioassays constitute a large number of experiments performed with similar protocols, and thus are ideal for this approach. The model chosen (2) to fit the data generated is defined as follows.

The measure of dose chosen is defined to be the lifetime integrated mass of material ingested measured in bodyweights. (See Appendix A for the methods used for computing doses). For conve-

*Energy and Environmental Policy Center, Jefferson Physics Laboratory, Harvard University, Cambridge, MA 02138.

nience this dose is expressed as an average dose rate by dividing by nominal lifetimes.

Tumor occurrence is a stochastic process occurring with a probability p related to the dose d defined above by

$$p = 1 - (1 - \alpha) \exp \{ - \beta d / (1 - \alpha) \} \quad \begin{array}{l} \alpha \geq 0 \\ \beta \geq 0 \end{array} \quad (1)$$

where α and β are the parameters of this model, α representing the background tumor incidence while β is a measure of the carcinogenic strength of the material under test. β is referred to as the potency of the material.

In any group of N animals fed the test material at a (group average) dose d , tumor incidence r is a random variable with a binomial probability distribution given by

$$P(r) = \binom{N}{r} p^r (1 - p)^{N - r} \quad (2)$$

where $P(r)$ is the probability of observing r animals with tumors, and p is given by the expression (1) above.

By using this definition of the model, an estimate of the potency β and approximate confidence limits associated with the uncertainties due to numbers of animals tested may be found using maximum likelihood methods (Appendix B). In addition, the significance of this estimated potency can be calculated exactly under the assumptions of the model (Appendix C).

This analysis has been repeated for almost every set of results reported in the summary tables set out in the NCI Carcinogenesis Bioassay Reports. Data from pooled or historic control groups were used in preference to data from the matched control groups, and time adjusted data were used preferentially whenever presented. Testicular tumors in male Fischer 344 rats were entirely omitted from the analysis.

In addition, similar calculations have been performed with a modified measure of dose, chosen so

as to weight the doses given at each age with a weight factor corresponding to that expected in a linearised Doll-Armitage model (3) in which the carcinogen affects just one stage (Appendix D). The essential results obtained do not differ (for the parameters so far tested) from those found with the unweighted measure of dose described above, and are not presented separately.

In order to compare between species it is necessary to select amongst all the potency estimates obtained. The available selection criteria, together with the choices used in this paper, are: (A) chemical: whole database of 187 chemicals; (B1) species: B6C3F1 mouse, Osborne Mendel rat, Fischer 344 rat; (B2) sex: male, female, either; (B3) tumor type/site: any; (C) significance of potency estimate: various (see Table 1), no correction made for multiple comparisons; (D) size of potency estimate: largest.

(B1), (B2) and (B3) were logically ANDed together. These criteria were applied in the order given to obtain a single value for each chemical tested. Confidence bounds were taken to be the lower confidence limit associated with the potency estimate selected and the highest upper confidence limit on all potency estimates satisfying (A)–(B3). Thus the upper confidence bound represents a measure of the sensitivity of the experiment. In most cases this upper confidence bound came from the same data that generated the selected potency estimate. By generating estimates of potency in this way with differing choices (of species, sex, tumor type/site) it is possible to compare them between these choices.

The purpose of this paper is to compare between species, so repeating the selections with differing species and all other selection criteria similar gives the required data. Each chemical may then be represented on a graph by a point with coordinates corresponding to the potency estimates obtained in the different species. Of course, for many chemicals there were no results significant in one or both species. Those with no significant result in either spe-

Table 1.

Species	Logarithm of potency ratio					
	$p = 0.01$		$p = 0.025$		$p = 0.05$	
	Mean \pm S.D.	n	Mean \pm S.D.	n	Mean \pm S.D.	n
(F344 or OM)/B6C3F1	0.283 \pm 0.716	36	0.207 \pm 0.681	51	0.170 \pm 0.662	66
F344/B6C3F1	0.456 \pm 0.669	26	0.362 \pm 0.648	37	0.337 \pm 0.609	48
OM/B6C3F1	-0.169 \pm 0.662	10	-0.201 \pm 0.610	14	-0.276 \pm 0.603	18
Males						
F344/B6C3F1	0.632 \pm 0.760	12	0.370 \pm 0.746	20	0.436 \pm 0.714	29
OM/B6C3F1	-0.050 \pm 0.670	7	-0.122 \pm 0.692	8	-0.300 \pm 0.769	12
Females						
F344/B6C3F1	0.369 \pm 0.631	16	0.376 \pm 0.579	20	0.280 \pm 0.569	28
OM/B6C3F1	0.004 \pm 0.645	8	0.012 \pm 0.631	12	-0.098 \pm 0.646	14

cies can be discarded as providing no useful information, but those with just one significant result are potential exceptions to any relationship found between potency estimates in the two species compared. In these cases one can plot the significant result in one species versus the upper confidence bound found for all results in the other species, and check whether this disagrees with any apparent interspecies relations. The confidence limits shown in this paper are approximate 95% limits (90% confidence intervals).

There are sufficient data to present comparisons between B6C3F1 mice and Fischer 344 (F344) rats, and between B6C3F1 mice and Osborne Mendel (OM) rats. The comparisons extend over nonoverlapping sets of chemicals, since each chemical was tested in only one strain of rat.

Results

Since the usual method of extrapolating between species (as used in regulatory efforts aimed at estimating effects on humans) is to take the largest significant effect seen in one species (independent of sex), we present results first with no sex discrimination. Figures 1, 2 and 3 are plots of the logarithm (to base 10) of potency in B6C3F1 mice versus the same in F344 rats. The criteria used (B and C above) are shown as labels on the graphs. A total of 78 chemicals showed a significant ($p \leq 0.025$, uncorrected for multiple comparisons) result in one or both of these species. The 37 significant in both are plotted in Fig-

ure 1, with the confidence limits, and again in Figure 2 without, while 22 significant in just one species are shown in Figure 3. The remaining 17 significant in just one species are not plotted, since they are obviously in agreement with the results of Figure 1 (i.e., on these plots, the arrows indicating the possible range of values point towards the dashed lines). The dashed line is a least-squares unit slope

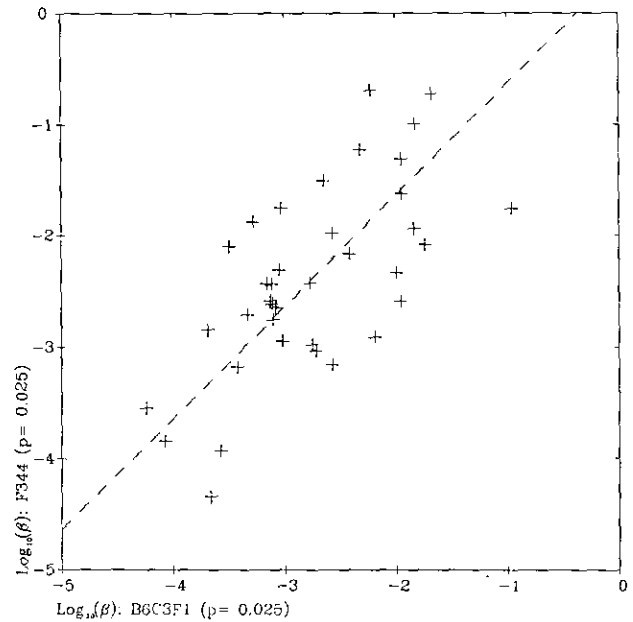


FIGURE 2. As Fig. 1 but without error bars.

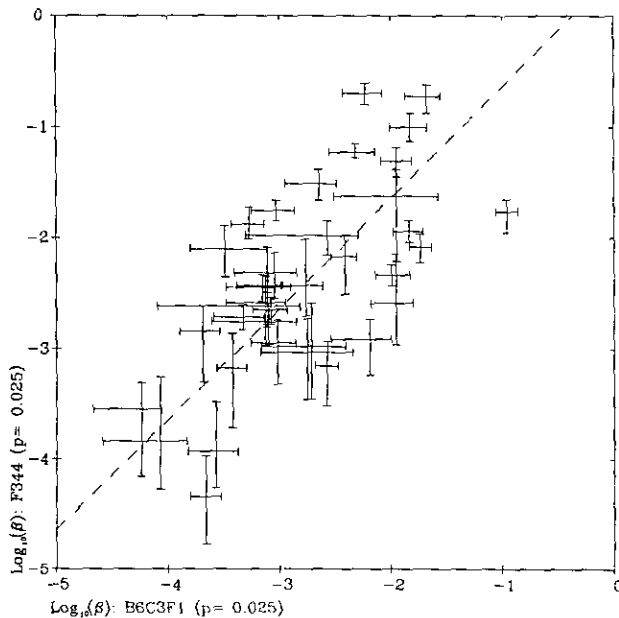


FIGURE 1. Logarithms of carcinogenic potency in B6C3F1 mice versus the same in F344 rats.

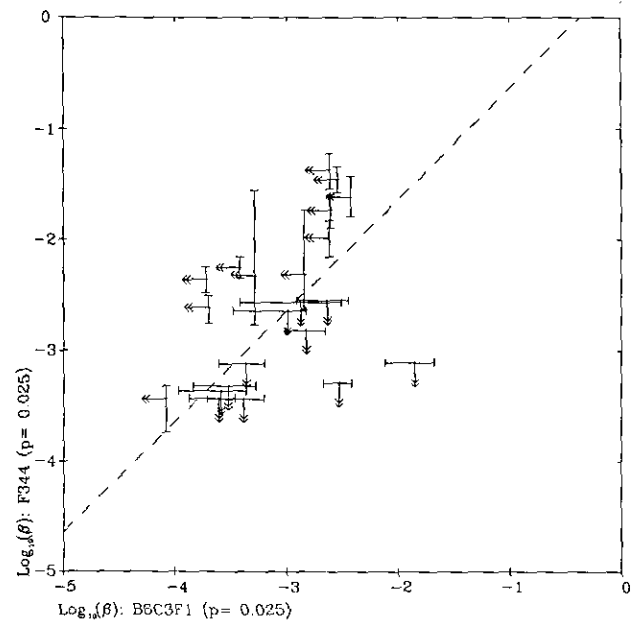


FIGURE 3. Cases which are not in self-evident agreement with the dashed line of Fig. 1.

regression on to the 37 points of Figure 1, a line corresponding to proportionality between species responses. Some statistics of these and later plots are shown in Table 1.

Figure 1 is included to illustrate that confidence intervals calculated solely on the basis of sample sizes and binomial distributions do not explain the variance between results in different species, most of which must thus arise elsewhere. The distribution of points is clearer in Figure 2, where the confidence intervals have been removed. Clearly, any attempt to extrapolate between these two species must take into account the uncertainty indicated by this graph. The standard deviation about the regression line (Table 1) is 0.66, corresponding to a factor of 4.5. We have elsewhere suggested (4) how this may be taken into account.

On Figure 3 are plotted those cases which do not naively agree with the regression line. As can be seen, most appear consistent with a distribution leading to Figures 1 and 2, with only one (piperonyl sulfoxide, causing hepatocellular carcinoma in male mice) deviating substantially. Figures 4 and 5 show similar results for B6C3F1 mice versus OM rats. Once again, one can note that confidence intervals based on numbers of animals tested fail to cover the range of deviations between species, but that within these larger deviations there appear to be no obvious exceptional cases—or rather that the experiments as designed have not shown up any such cases.

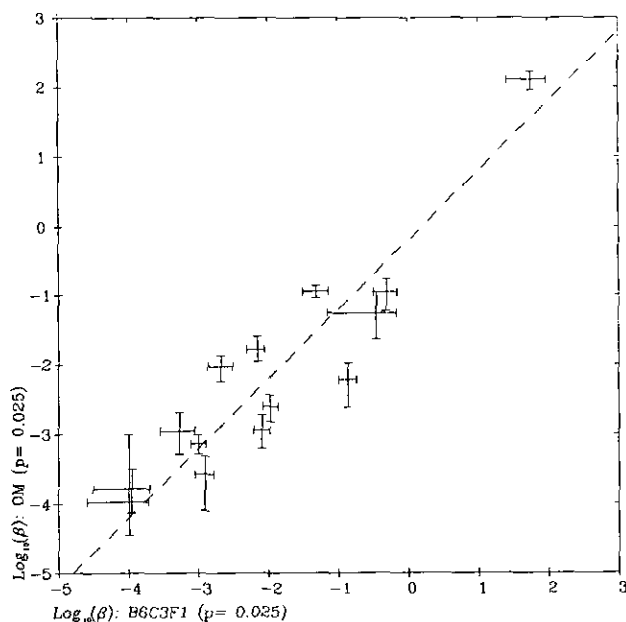


FIGURE 4. Logarithms of carcinogenic potency in B6C3F1 mice versus the same in OM rats.

This process can be repeated for various selections. The general features of the resulting plots are similar to those displayed, and a few of the statistics generated are shown in Table 1. The numerical values shown in Table 1 correspond to the logarithmic plots displayed, and so refer to the logarithm (to base 10) of the ratios of potencies estimated in the species/strains in the left hand column. The values obtained are fairly stable against changes in the significance selection criterion, as may be seen by comparing the values given for $p \leq 0.01$, $p \leq 0.025$ and $p \leq 0.05$. All values agree within their estimated standard errors.

Discussion

The results illustrated here indicate that using this model, extrapolation of carcinogenic potency between rats and mice is uncertain, the standard deviation of the logarithm (to base 10) of the ratio of potencies measured in mice and rats amounting to about 0.65 (a factor of 4.5). Strictly speaking, the uncertainties found in Table 1 apply only to the dose-response model as defined earlier in this paper, and then only for high doses. However, it seems unlikely that any other dose-response model is likely to do much better in this high dose regime, except possibly if it is highly nonlinear in the range of doses tested. This possibility is, of course, open to test in the way outlined in this paper.

As mentioned earlier, changing the measure of

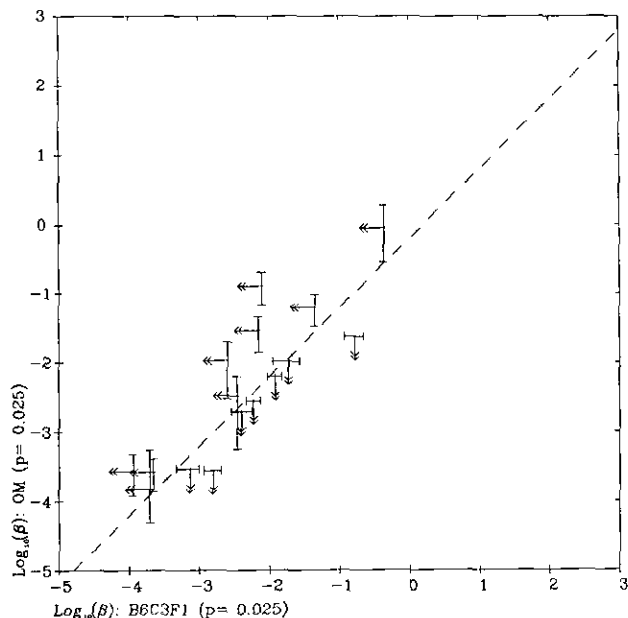


FIGURE 5. Cases which are not in self-evident agreement with the dashed line of Fig. 4.

Table 2. Based on Table 1 $p = 0.025$ data.

Species	Logarithm of potency ratios	
	Body weight basis	Surface area basis ^a
(F344 or OM)/B6C3F1	0.21 ± 0.10	-0.17 ± 0.10
F344/B6C3F1	0.36 ± 0.11	0.03 ± 0.11
OM/B6C3F1	-0.20 ± 0.16	-0.61 ± 0.16
Males		
F344/B6C3F1	0.37 ± 0.17	0.02 ± 0.17
OM/B6C3F1	-0.12 ± 0.24	-0.54 ± 0.24
Females		
F344/B6C3F1	0.38 ± 0.13	0.06 ± 0.13
OM/B6C3F1	0.01 ± 0.18	-0.37 ± 0.18

^a Computed by assuming surface area is proportional to the two-thirds power of body weight, with the following typical body weights used: B6C3F1 (m) 32g, B6C3F1 (f) 28; OM (m) 600g, OM (f) 400g; F344 (m) 350g, F344 (f) 250g.

dose by weighting dose rates with their time of administration has little effect on this measure of uncertainty. This result is not surprising, since many of the chemicals contributing to the large variance were fed to the animals at constant dose rates in these experiments, so that this form of weighting has little effect on the ratios of doses fed to the different animals, no matter how those doses are measured. Similarly, using a dose measure based on surface area (e.g., mg/m²-day) will not affect his measure of uncertainty, since such a dose measure is related to the one used (mg/kg-day) by a surface area to weight ratio which has much lower variance within a species. The only effect of changing to such a dose measure is to change the mean value of the interspecies potency ratio, represented by the position of the intercept of the dashed lines on the graphs shown.

The data of Table 1 are equivocal in their support for extrapolations between species solely on the basis of dose measured per unit surface area (practically, two-thirds power of body weight). Suppose that β_1 and β_2 are potencies measured in species 1 and 2 on the basis of doses d_1 and d_2 measured in mass per unit body weight, and that similar, primed, variables representing similar quantities are defined for doses measured in mass per unit surface area.

Then $\beta'/\beta = d/d' \simeq W^{-1/3}$, where W is body weight, and so $\beta_1'/\beta_2' = (\beta_1/\beta_2) (W_1/W_2)^{-1/3}$. Using typical weights for the animals in these experiments gives the results shown in Table 2. As can be seen, extrapolation on a bodyweight basis appears best for the OM to B6C3F1 comparison, on a surface area basis for the F344 to B6C3F1 comparison, and neither is particularly good if all the data are combined for a single rat to mouse comparison. Although it has been shown (5) that an acute toxicity measure may be extrapolated between species by using measure of dose based on surface area (with an uncertainty substantially less than those ob-

served in this paper), the rat appeared to be exceptional in giving biased estimates using this procedure. We are left wondering if the rat is exceptional for carcinogenicity extrapolation also.

Appendix A: Dose Calculations

Dose Administered on Weight Basis. For those experiments in which the material was administered by gavage, doses were specified and administered on a weight basis (mg/kg), usually by adjusting dose sizes according to average weights of groups of animals. Total lifetime dose was obtained by summing over all individual doses, and normalised to a dose rate (mg/kg-day) by dividing by nominal lifetimes, assumed to be 91 weeks for mice and 104 weeks for rats.

Dose Administered as a Fraction of Diet. Here doses were specified as a fraction (ppm or %) of the diet. To compute a lifetime integrated dose in body weights required the use of feeding curves for the experimental animals. Data on feeding of OM rats and B6C3F1 mice were provided in Carcinogenesis Bioassay Report No. TR2. From these data it was found that the cumulative intake of food measured in body weights (obtained by dividing weekly or four weekly food intake by the mean of the weights of the animals at the beginning and end of each week or four weeks, and summing from the beginning of the experiment) are well fitted by a curve of the form:

$$F(t) = b(1 - \exp\{-at\}) + dt + ct^2$$

where $F(t)$ is the cumulative intake in body weights at age t and a, b, c, d are parameters estimated from the data.

The parameters were estimated from data on groups of animals on control diets. To allow for weight differences between different groups of animals, and between OM and F344 rats, it was as-

sumed that food intake is proportional to the two-thirds power of body weight, so that food intake measured in body weights is proportional to the inverse one-third power of body weight. Thus the cumulative material intake due to a concentration C of material in the diet a fraction K of the time from ages t_1 to t_2 could be expressed as:

$$KCM^{-1/3}[F(t_2) - F(t_1)]$$

for an animal of full grown weight M . Scaling the parameters a, b, c, d allowed this expression to directly give the average dose rate (mg/kg-day), using nominal lifetimes of 91 weeks for mice, 104 weeks for rats. Full grown weight was estimated by eye from the growth curves of each experiment, and corresponded to weight at 60-70 weeks of age. The fraction K was used when dosing was only carried out for some fraction of the week. Adding up expressions like this over the concentration-age schedule for the experiment gave the total average dose rate.

The values of the parameters used (with age measured in weeks, concentration in ppm, and weight in grams) are given in Table A-1.

Table A-1

	a	b	c	d
Male rats	0.132933	8.77459E-2	0.0	3.00191E-3
Female rats	0.184978	0.112783	0.0	3.45399E-3
Male mice	5.90133E-2	8.89311E-2	1.60808E-5	1.84592E-3
Female mice	7.92837E-2	1.19209E-1	8.32436E-6	2.58261E-3

Dose Administered in Water. A simple approach was taken, scaling water intake with full grown body weight and normalizing to standard animals. Fitting the scaling law to standard mouse, rat, dog and man gave water intake as:

$$T = 0.3629 W^{-0.2125}$$

where T is daily water intake in body weights and W is body weight in grams. This relation was assumed to hold for all ages, so that simply multiplying the time integrated concentration of material in water (in ppm-days) and dividing by nominal lifetimes (in days) gave average dose rates in mg/kg-day.

Appendix B: Estimation of Potency and Confidence Limits on Potency

Suppose an experiment is performed in k groups of animals, with results: number of animals in group i : N_i ; average dose of group i : d_i ; and number of ani-

mals with tumors in group i : r_i . The expected probability of tumors in this group is given by:

$$p_i = 1 - (1 - \alpha) \exp\{-\beta d_i / (1 - \alpha)\}$$

where the parameters α and β are unknown. The likelihood function for the k experiments is then

$$L(\alpha, \beta) = \prod_{i=1}^k \binom{N_i}{r_i} p_i^{r_i} (1 - p_i)^{N_i - r_i}$$

Maximizing this with respect to α and β , subject to $\alpha \geq 0$ and $\beta \geq 0$ gives estimates $\hat{\alpha}$ and $\hat{\beta}$ of the parameters.

Approximate $(1 - 2e)$ confidence limits are obtained by finding the two solutions β_e of:

$$2\{\ln[L(\hat{\alpha}, \hat{\beta})] - \ln[L(\alpha_e, \beta_e)]\} = \chi(1 - 2e)$$

where $\chi(1 - 2e)$ is the cumulative $(1 - 2e)$ point of the chi-squared distribution with one degree of freedom and α_e is the value of α which maximizes $L(\alpha, \beta_e)$.

Appendix C: Significance of Potency Estimate

The significance of the potency estimate is the probability of observing a potency estimate as high or higher than the one actually observed, under the null hypothesis that the true value of potency is zero, or equivalently that all groups of experimental animals are samples from the same population. For any experimental design represented by:

$$\left. \begin{array}{l} \text{groups of } n_i \text{ animals} \\ \text{dosed at levels } d_i \end{array} \right\} \quad i = 1, 2, \dots, k$$

we can obtain a potency estimate $\hat{\beta}$ which is a function of the number of animals with tumors in each group, say s_i ; $i = 1, 2, \dots, k$, so

$$\hat{\beta} = \hat{\beta}(s_1, s_2, \dots, s_k)$$

The observed result is then $\hat{\beta} = \hat{\beta}(r_1, r_2, \dots, r_k)$ where r_i is the observed number of tumorous animals in group i . Under the null hypothesis, all groups are equivalent, so the

$$R = \sum_{i=1}^k r_i$$

tumorous animals, are equally likely to be anywhere among the

$$N = \sum_{i=1}^k n_i$$

animals.

Hence the probability of observing an estimated potency as large as actually seen is just the probability of getting the tumorous animals distributed

into the k groups in such a way that the potency estimate is as large or larger than β , that is:

$$p = \sum \frac{\prod_{i=1}^k \binom{n_i}{s_i}}{\binom{N}{R}}$$

where the sum is over all values of s_1, s_2, \dots, s_k which satisfy

$$\sum_{i=1}^k s_i = R$$

and

$$\hat{\beta}(s_1, s_2, \dots, s_k) \geq \hat{\beta} = \hat{\beta}(r_1, r_2, \dots, r_k)$$

Appendix D: Dose Weighting for a Linearized Doll-Armitage Model

In this model it is assumed that a tumor arises after a sequence of M events have occurred in order. Assume the probability per unit time of the j th event is $p_j(t)$ at age t . Then if $P_n(t)$ is the probability that n events have occurred by age t , we have:

$$dx_n/dt = (1 - \exp\{-x_{n-1}\})p_n \quad n > 1$$

where $x_n = -\ln(1 - P_n)$

$$dx_1/dt = p_1$$

This may be linearized by assuming all the x to be small, in which case

$$dx_n/dt = x_{n-1} p_n \quad n > 1$$

Then we get

$$x_m(t) =$$

$$\int_0^t dt_m p_m(t_m) \int_0^{t_m} dt_{m-1} p_{m-1}(t_{m-1}) \dots \int_0^{t_2} dt_1 p_1(t_1)$$

If all the $p_i(t)$ are constants except for $p_j(t)$, ($j \leq m$), this becomes

$$x_m(t) =$$

$$\left\{ \prod_{\substack{i=1 \\ i \neq j}}^m p_i \right\} \frac{1}{(m-j)!(j-1)!} \int_0^t dx (t-x)^{m-j} x^{j-1} p_j(x)$$

If p_j depends linearly on dose rate $d(t)$, then this can be written in the form

$$x_m(t) = A + B \int_0^t dx (t-x)^{m-j} x^{j-1} d(x)$$

Setting t to be the length of a lifetime and expressing P_m in terms of x_m allows writing P_m in the form of the model adopted in this paper, with a dose defined by

$$D = (1/L_0^m) \int_0^L dx (L-x)^{m-j} x^{j-1} d(x)$$

where $d(t)$ is the dose rate at age t , L is the actual lifetime and L_0 is a nominal lifetime.

This form of weighting factor was used to determine a dose variable, where the dose rate was measured in bodyweights per day and computed as in Appendix A. The integration was performed by summing values computed daily.

This work was performed under DOE contract number DE-AC02-81EV10598. I should like to thank P. Fraenkel for writing initial versions of several of the computer programs used.

REFERENCES

1. Purchase, I. H. F. Interspecies comparisons of carcinogenicity. *Brit. J. Cancer* 41: 454-468 (1980).
2. Crouch, E., and Wilson, R. Interspecies comparison of carcinogenic potency. *J. Toxicol. Environ. Health* 5: 1095-1118 (1979).
3. Armitage, P., and Doll, R. The age distribution of cancer and a multistage theory of carcinogenesis. *Brit. J. Cancer* 8: 1-12 (1954).
4. Crouch, E., and Wilson, R. Regulation of carcinogens. *Risk Anal.* 1: 47-57 (1981).
5. Freireich, E. J., Gehan, E. A., Rall, D. P., Schmidt, L. H., and Skipper, H. E. Quantitative comparison of toxicity of anticancer agents in mouse, rat, hamster, dog, monkey, and man. *Cancer Chemother. Rep.* 50: 219-244 (1966).