

The TD₅₀: A Proposed General Convention for the Numerical Description of the Carcinogenic Potency of Chemicals in Chronic-Exposure Animal Experiments

by Richard Peto,* Malcolm C. Pike,*† Leslie Bernstein,†
Lois Swirsky Gold‡ and Bruce N. Ames**

A generally accepted format for the numerical description of the carcinogenic potency of a particular chemical in a particular strain of animals is desirable so that statements from different sources about potency and attempts by different authors to correlate potency with particular laboratory measurements will be comparable. The choice of an appropriate standard format is to a certain extent arbitrary. In this paper we recommend that the TD₅₀ (tumorigenic dose rate 50) be used. TD₅₀ can be calculated for a single target site or combination of sites. The TD₅₀, in analogy with the LD₅₀, is defined as that chronic dose rate (in mg/kg body weight/day) which would halve the actuarially adjusted percentage of tumor-free animals at the end of a standard experiment time—the “standard lifespan” for the species. This paper consists of a brief discussion of the TD₅₀, sufficient to make the general reader familiar with the properties of such an index, an appendix discussing methods for its estimation and certain conventions we have adopted for use in analyzing “nonstandard” experiments. A major problem in calculating any index of carcinogenic potency is that much published material gives only the final crude percentage of tumor-bearing animals at each dose, instead of percentages adjusted for the effects of intercurrent mortality or data from which these adjusted percentages can be derived. If the dose level administered to the animals is toxic, then premature death from nonneoplastic causes may prevent some dosed animals that would have developed tumors from actually doing so. This will particularly affect the high-dose group. Consequently, any estimate of carcinogenic potency that is based on crude percentages of tumor-bearing animals, not adjusted for intercurrent mortality, may underestimate the carcinogenicity of the test material.

A Numerical Index: the TD₅₀

For various purposes, such as a quantitative comparison of the results from carcinogenicity experiments with results from various *in vitro* tests, a numerical index of carcinogenic potency must be estimated from animal carcinogenicity data. Many *ad hoc* indices can be proposed, none clearly superior to all others. In this situation, it would clearly be desirable if most workers would generally adopt the same index, and we propose that, by convention, a particular numerical measure of

carcinogenic potency (the TD₅₀) be adopted as this standard index.

In classical acute toxicology, the LD₅₀ (lethal dose 50) of a chemical is defined as that dose of the chemical which kills 50% of the test animals. A large value for the LD₅₀ indicates, of course, a substance of low acute toxicity, while a small LD₅₀ indicates a very potent poison. Although the LD₅₀ varies with strain, species and experimental conditions, it has proved to be a useful and practical measure of acute toxicity, and is widely used and understood. In order to adopt some roughly analogous measure for the tumorigenicity of a particular agent, the TD₅₀ will be defined as the tumorigenic dose rate for 50% of the test animals, i.e., for a given target site(s), the TD₅₀ is that chronic dose rate (in mg/kg body weight/day) which would give half of the animals tumors within some standard experiment time—the “standard lifespan” for the species. There are, however, three points that must be clarified before such a definition of the TD₅₀ is possible.

*I.C.R.F. Cancer Studies Unit, Nuffield Department of Clinical Medicine, Radcliffe Infirmary, Oxford OX2 6HE, United Kingdom.

†Department of Preventive Medicine, University of Southern California School of Medicine, 2025 Zonal Avenue, Los Angeles, CA 90033.

‡Biology and Medicine Division, Lawrence Berkeley Laboratory, Berkeley, CA 94720. Author to whom correspondence should be addressed in the U.S.

**Department of Biochemistry, University of California, Berkeley, CA 94720.

First, how is the TD_{50} to be defined if some of the control animals also get tumors? The most natural definition to adopt seems to be that the TD_{50} is that dose rate which halves the probability of remaining tumorless to the end of the standard lifespan. Using this definition, if 20% of the controls get the tumor(s) of interest, then the TD_{50} will be the dose rate that leads to a total of 60% of the treated animals developing that tumor. This definition is analogous to that proposed for LD_{50} when deaths occur in the the control group (1,2). It obviously makes good sense when the spontaneous and treatment-related tumors arise by independent mechanisms, but it does not implicitly assume that this is the case.

Second, some tumors are observed in a "fatal" context (i.e., they are the direct or indirect cause of the death of their host), while others are observed only in an "incidental" context (i.e., they do not cause the death of their host, and are discovered only because the host has been sacrificed, has died of nonneoplastic treatment toxicity, or has died of some unrelated disease). By convention, in defining the TD_{50} , we shall count all animals bearing the tumor of interest, irrespective of whether their tumors were observed in a fatal or in an incidental context. We have discussed at length elsewhere the statistical relevance of the distinction between incidental and fatal tumors (3). Here, however, we are concerned only with the biological question of whether, after 30% of a group of animals have died as a result of a given type of tumor and a further 30% of the group are already carrying such tumors but are still alive, we define the proportion of tumorless animals to be 70% or 40%. To us the latter seems preferable, since otherwise we would be forced to ignore all of the information that is found when an experiment is terminated and the survivors are autopsied. When substantial numbers of terminally sacrificed animals have internal tumors that are detectable only at autopsy, the "incidental" tumors found may well provide more information than all of the earlier findings put together. Moreover, to descend from the theoretical to the practical, the information as to the context in which tumors are observed is, with the exception of terminal sacrifice, rarely recorded or published.

Third, most types of cancer are typically diseases of old age, and premature deaths from other causes will prevent some cases of cancer that would have arisen in old age from actually doing so. Intercurrent mortality is a moderate nuisance in experiments with "nontoxic" doses of carcinogens (i.e., carcinogens administered at dose levels that do not materially affect nonneoplastic causes of death), and may be a serious nuisance in experiments with toxic carcinogens, where more high-dose animals than controls die prematurely from nonneoplastic causes. Papers reporting that the crude percentage of animals which develop tumors was decreased at high dose(s) are probably reporting artifacts that are due to failure to make proper allowance for the effects of competing risks on tumor yields (3). The definition of the TD_{50} should therefore be based not on the crude

percentages of animals that develop tumors, but rather on some estimate of the percentages of animals that would have done so had intercurrent mortality been prevented.

There are various ways in which these mortality-corrected probabilities may be estimated. Ideally, one might first estimate the proportion of animals that would still be alive at the end of the standard lifespan if all causes of death other than the tumor type(s) of interest were prevented, then estimate the proportion of the survivors at the end of the standard lifespan that are free of occult tumors of the type(s) of interest, and finally multiply these two proportions together. In practice, few, if any, experiments provide the full data that are needed to achieve this. If such detailed time-to-tumor data are available, the least unsatisfactory approximation in many instances seems to be to use "actuarial," or "death-rate" methods (3) on all tumors. These methods are exactly appropriate for premature deaths as a result of the tumor type of interest and are algebraically equivalent to prevalence methods for all tumors found at terminal sacrifice, but lead to some overestimation of the final tumor yield when there are many premature deaths at which incidental tumors are found. In some instances, however, especially when the tumors that arise are principally lesions that are considered unlikely to have been fatal, "prevalence-rate" methods (3) for all neoplasms may instead be preferred. In principle, an appropriate combination of death-rate and prevalence-rate methods can be used for TD_{50} estimation; in practice, this should be done only when sufficiently extensive data on prevalence rates are available for gross numerical instabilities to be avoided.

Although we give details in the Appendix of certain statistical procedures we have adopted for estimating the TD_{50} from experimental data in our Carcinogenic Potency Database described in the accompanying paper (4), these procedures are not intended to be part of its definition and other statistical procedures may be employed.

Definition

For any particular sex, strain, species and set of experimental conditions, the TD_{50} is the dose rate (in mg/kg body weight/day) that, if administered chronically for a standard period—the "standard lifespan" of the species—will halve the mortality-corrected estimate of the probability of remaining tumorless throughout that period.

As defined, a TD_{50} can be computed either for a particular category of neoplastic lesion (e.g. "malignant tumors only," "liver tumors only," etc.) or for all tumors. There is no need for absolute uniformity in this, for on different occasions, different tumor categories will be of interest. In the absence of any special overriding considerations, we propose that the category studied should be either "those tumor types that are strongly affected by treatment" or "all tumor types, benign or

malignant.* If treatment strongly affects some tumor types and has no material effect on any others, then use of either of these two categories will usually yield approximately the same TD_{50} . When estimating a TD_{50} from published reports that do not give details of the exact category of tumors one would like to study, it is necessary to choose from those tumor categories that are adequately documented. It may then be best to study "the category of tumors that are statistically significantly related to treatment," despite the bias which this can in principle cause. In our analyses of chemicals tested in the Carcinogenesis Bioassay Program of the National Cancer Institute/National Toxicology Program (NCI/NTP) given in the accompanying paper, we have generally utilized analyses of individual categories of tumors considered to be treatment-related in the NCI/NTP reports, the aggregate of all sites considered to be treatment-related in the NCI/NTP reports, and all tumor-bearing animals (TBAs) excluding only testis tumors in F344 rats. For the general published literature we have used TBAs and individual sites, but lack the data to aggregate the treatment-related sites.

Why the TD_{50} Rather Than Some Other Index of Potency?

Indices such as the TD_{10} or TD_{90} can be defined and are probably as good as the TD_{50} if they can be as reliably estimated. One advantage of the TD_{50} is that the experimental dose range will often include it, which makes for statistically accurate estimation. Also, there is the useful analogy with the LD_{50} .

An alternative to the TD_{50} might be the "doubling dose" for some category of tumors, i.e., the dose rate needed to double the spontaneous tumor rate. The advantage of this index is that it might on theoretical grounds, although not on any direct evidence, be expected to be more reproducible in different species. The overwhelming disadvantage of using a doubling dose to characterize experimental data is that the spontaneous frequency of most tumors is difficult to characterize accurately, and so, estimated doubling doses would be subject to severe sampling errors.

Indices Proposed by Other Authors

Meselson and Russell (5) defined an index of carcinogenic potency, which in the absence of intercurrent mortality is equivalent to $\ln 2/TD_{50}$ (i.e., to about $0.693/TD_{50}$). The description of the method of calculating the index given in their paper clearly shows that they adjust for intercurrent mortality, when this is

possible, but it is unclear how they propose allowing for the percentage of tumors in the control groups.

Crouch and Wilson (6) proposed an index of potency which will in practice often be closely correlated with the TD_{50} , but where noteworthy discrepancies do exist, the TD_{50} is preferable. Formally, Crouch and Wilson assumed that $-\ln(1-P) = a + bd$, P being the proportion developing tumors at dose rate d and a and b being constants to be estimated from the data. If this model fitted the data and P were corrected for intercurrent mortality, Meselson and Russell (5) would estimate the carcinogenic potency as b , and we would estimate the TD_{50} as $0.693/b$. Crouch and Wilson, however, ignore the possibility of correction for intercurrent mortality, and define potency not as b but as $b \exp[-a]$. This latter convention is in principle somewhat unsatisfactory because it is systematically biased by the occurrence of spontaneous tumors by mechanisms unrelated to the carcinogenic effects of the test agent. The bias is by a factor of $\exp[-a]$ so that in practice it is important only when the spontaneous frequency, $1 - \exp[-a]$, of tumors is large.

Finally, the National Academy of Sciences' Committee on Prototype Explicit Analyses for Pesticides (7) defined a Carcinogenic Activity Indicator (CAI) for each dose group in a particular experiment by $CAI = \text{"percent response"} \div \text{dose}$, where the "percent response" is the difference between the mortality-corrected (life-table) percentages of tumor-bearing animals in the treated and control groups. The CAIs calculated from each treatment group will usually be different and these differences will depend systematically on the dose level that was tested. (Even if they do not do so at low doses they must do so at high doses, since the excess percentage affected cannot exceed 100%.) This makes the CAI somewhat unsatisfactory. These problems could be avoided by specifying that the CAI to be used for a chemical is the CAI at that dose which gives half the animals tumors, or halves the probability of remaining tumorless. The CAI would then fairly closely resemble the TD_{50} , except that in the proposed definition of the CAI no explicit account was taken of the duration of the study. The CAI, like Crouch and Wilson's potency index, is also systematically biased by variations in the frequency of spontaneous tumors. Thus, although the CAI will generally be quite closely correlated with the TD_{50} , where discrepancies arise the TD_{50} is preferable.

Confidence Intervals

The TD_{50} estimated from a particular experiment is subject to the usual statistical uncertainties, and it is usually useful to estimate a confidence interval about it. Invariably, there will also be nonstatistical uncertainties in the conduct of any experiment, and this suggests that a wide statistical confidence interval will be more "realistic"; we recommend that a 99% confidence coefficient be used rather than the more usual 95%. In an experiment where the statistical significance of TD_{50} is $p < 0.01$ (two-tailed), the confidence interval will be

*Except, perhaps, for one or a few prespecified types of tumor that have such a high spontaneous frequency that they would swamp the data (e.g., interstitial-cell testis tumors in male Fischer F344 rats, a neoplasm that eventually affects almost all such animals).

two-sided: for example, a 99% confidence interval of 1.5 to 6.7 mg/kg body weight/day for the TD_{50} would suggest that (in the absence of other causes of death) 1.5 mg/kg body weight/day would not halve the proportion of tumorless survivors at the end of the standard lifespan, while 6.7 mg/kg body weight/day would more than halve it.

In an experiment where the statistical significance of TD_{50} is $p > 0.01$ (two-tailed), the confidence interval for the TD_{50} will be open at one end; for example, a confidence interval with a lower limit of 10 mg/kg body weight/day for the TD_{50} and no upper limit suggests that 10 mg/kg body weight/day would probably (in the absence of other causes of death) fail to halve the proportion of tumorless survivors, but makes no definite statement about the carcinogenic effects of higher doses. The results from a statistically nonsignificant ($p > 0.01$) experiment thus give a lower limit to the TD_{50} , but do not demonstrate noncarcinogenicity. A fuller discussion of the meaning and proper uses of p -values is given elsewhere (3).

TD_{50} Estimated, with Bias, from Unadjusted Percentages of Tumor-Bearing Animals

In chronic carcinogenicity studies, the cancers that arise often do so quite late in the lifespan of the animals concerned. As has already been noted, the number of animals that eventually develop cancer at any one particular age therefore depends on the number of animals that survive to that age (which may be reduced by intercurrent mortality to a different extent in each treatment group), as well as on the proportion of these survivors that will then develop new tumors in the near future. "Actuarial" (life-table) analysis overcomes the gross effects of mortality[†] from other causes by estimating the number of animals that would have developed tumors if, hypothetically, all nonneoplastic deaths had been prevented. This analysis is the method that we have used in our Carcinogenic Potency Database when time-to-tumor data are available.

In many positive carcinogenicity bioassays, the non-neoplastic toxicity of the highest administered dose causes sufficient early deaths to reduce quite substantially the proportion of animals that survive long enough to develop cancer. On occasion, this effect is so extreme that the proportion of animals that get cancer before they die is actually lower in the group given the highest dose of carcinogen than in lower dose groups.

[†]As already noted, precise correction requires that a distinction be made between tumors observed in an incidental and in a fatal context, and this information is not generally available in present-day experimental reports, let alone in the past literature. Although the inappropriate use of actuarial methods for incidental neoplasms may somewhat overcorrect for the effects of intercurrent mortality on tumor yields, it may nevertheless be the best technique to use in practice (see above and the appendix).

Although there is likely to be a fairly direct relationship between carcinogenicity and the actuarially adjusted percentages of tumor-bearing animals, the relationship between carcinogenicity and the crude, unadjusted, percentages of tumor-bearing animals may thus be distorted or even inverted. Estimation of the TD_{50} (or any other numerical potency index) from crude unadjusted data may therefore underestimate the carcinogenicity of the test material. Although confidence intervals can be derived for the effects of random variation on a TD_{50} estimated from crude percentages, the systematic bias caused by intercurrent mortality cannot usually be quantified. If estimated from data on crude unadjusted percentages, the TD_{50} and its confidence limits will tend to be too high, underestimating the true potency of the chemical.

Whether or not authors of future carcinogenicity studies choose to estimate and to include in their published reports TD_{50} values (with confidence intervals) for the substances they study, it is important that they publish their data in sufficient detail to permit others to make allowance for intercurrent mortality. Lack of such data is a serious impediment to the accurate interpretation of many past studies (3).

Conclusion

The accompanying paper (4) provides a graphic display of the TD_{50} values (and associated confidence intervals) for all carcinogenicity tests reported before mid-1980 by the NCI/NTP Bioassay Program and for the long-term experiments from the published literature through mid-1981 that meet our standard criteria for selecting "suitable" tests.

As cancer research progresses, it will presumably be possible to account for more and more of the unexplained differences between the TD_{50} values observed in different carcinogenicity studies. Whether or not much immediate progress is possible, we hope that the plot of TD_{50} values which is provided in the accompanying paper will provoke more interest in the quantitative aspects of chemical carcinogenesis than currently exists. If nothing else, the database emphasizes how enormous the range of potency is; some weak carcinogens have a TD_{50} of over 1000 mg/kg body weight/day while others have a TD_{50} of under 0.001 mg/kg body weight/day. This millionfold range of potency is the context in which the tenfold differences which are sometimes observed between the potency of the same agent in different rodent species should be viewed. It is also the context in which different statistical approaches to estimating TD_{50} values should be viewed.

An obvious question to address with this list of TD_{50} values is to what extent these values are predictable from short-term test results. Of course, there are many qualitatively different carcinogens, and no single short-term test is likely to assess all of these different types of agent. For example, it is obvious that a mutagenicity test can be expected to indicate the carcinogenic haz-

ards only of chemicals that are carcinogenic chiefly by virtue of their mutagenicity, or by virtue of some closely allied property. A mutagenicity test is, therefore, unlikely to be of much assistance in assessing the carcinogenic effects of such agents as asbestos or phorbol esters or hormone replacement therapy, and will have nothing to say regarding such important factors as obesity or delayed maternity. Indeed, no single *in vitro* test is ever likely to rank a series of qualitatively different determinants of cancer rates in order of human importance of agents, although within one class (e.g., some subset of the mutagens, perhaps) it might be possible to develop tests which usefully predict at least the orders of magnitude of their quantitative carcinogenicity.

Appendix

Not all experiments involve continuous daily exposure throughout the test animal's lifetime, and many experiments do not continue for the full standard lifespan. In addition, few published papers report sufficient details on tumor occurrence to allow proper actuarial analysis. We have therefore had to adopt conventions regarding the estimation of the TD_{50} from "nonstandard" data. These conventions are described in this appendix, together with numerical algorithms for their implementation. There is, of course, little point in the general reader examining the appendix in detail, and no part of it is intended as part of the definition of the TD_{50} .

The definition of the TD_{50} is "the dose rate (in mg/kg body weight/day that, if administered chronically for a standard period—the 'standard lifespan' of the species—will halve the mortality-corrected estimate of the probability of remaining tumorless throughout that period." Quite apart from the need to define how long a "standard lifespan" is, this definition of the TD_{50} is of little practical use until we consider how parts per million are to be converted to units of mg/kg body weight/day, what to do when the experiment does not continue for the full standard lifespan, what to do when all the animals, including the controls, get tumors spontaneously, and how to generalize the definition to cover substances (like saccharin) where no practicable dose can possibly give half the animals cancer.

Standard Lifespan

By convention, most of the chronic studies for rats and mice currently conducted by the NCI/NTP are begun when the test animals are 6 to 8 weeks of age and are terminated after 90 to 110 weeks on test; the survivors are then killed and autopsied. In interpreting data from the large number of careful studies that have been published in the literature, it is convenient to adopt an experiment time in the range of 90 to 110 weeks as the conventional "lifespan" for rats and mice. For simplicity, we have adopted 2 years (104 weeks), a value that appears frequently in the literature, as the

conventional lifespan for rats and mice in our Carcinogenic Potency Database. Based on values given in the literature, we have adopted 2 years as the conventional lifespan for hamsters, 11 years as that for dogs and 20 years as that for rhesus, cynomolgous, and African green monkeys.

The TD_{50} for rats and mice is thus the dose rate in mg/kg body weight/day that, if administered for 2 years, will halve the actuarially adjusted proportion of tumorless survivors at that time. *Ad hoc* methods will later be suggested for estimating this TD_{50} from experiments that were terminated before or after 2 years. (Note that it would not suffice to consider only those tumors which were detected in the first 104 weeks of a 110 week experiment, because then one would miss the "incidental" tumors that were present at week 104 but not detected until later.)

Estimation of Mean (Lifelong) mg/kg Body Weight/day

With regard to TD_{50} , there are not at present sufficient data to determine whether mg/kg body weight/day, mg/m^2 surface area/day, mg/day, mg/lifetime, or ppm in the food or water will be the measure of dose that is most consistent among different species of laboratory animal, or between short-lived and long-lived species such as mice and men. The decision to define the TD_{50} in terms of mg/kg body weight/day is therefore largely arbitrary, and may require revision in the future when more data become available on quantitative interspecies differences in carcinogenicity. If, for the present, mg/kg body weight/day is chosen provisionally, conventions must be established to convert ppm in food or water into mg/kg body weight/day, and to deal with experiments where active treatment is discontinued some time before the experiment is terminated.

Even for a single experiment, there is no constant factor that exactly converts ppm in food or water to mg/kg body weight/day, because both food intake and body weight vary with age (and, in some experiments, with treatment). However, by assuming 100% absorption and adopting a set of standard values for each sex/species group which includes factors for daily food, water and air intake and average weight, we convert dose to mg/kg body weight/day. We know that these conversion factors will not be exactly correct, but, because their derivation considered weights and water intakes found in several sources in the literature, they are unlikely to be substantially in error. Details of these conversion factors are given in the following paper (4).

In some experiments, treatment is stopped before the scheduled end of the experiment, and the mean dose rate over the whole experimental period is therefore lower than the dose rate given during active treatment. For example, animals that receive 180 ppm in their food for 15 months in an experiment scheduled to end after 18 months could be considered to receive approximately an equivalent treatment to animals that

received 150 ppm ($180 \times 15/18$) ppm over the whole experiment period. This convention is probably as good as any, unless most of the treated animals developed tumors before treatment ceased, and we have therefore adopted it in the Carcinogenic Potency Database.

Correction for Experiments Which Terminate Prior to or After the Standard Lifespan

In an experiment which is terminated before the standard lifespan, the numbers of tumors found will be reduced, and the dose rate d needed to halve the proportion of tumorless animals at the end of the reduced period of observation will then be greater than the true TD_{50} . For this reason, one might estimate the true TD_{50} as fd , f^2d or f^3d , where we define $f = (\text{duration of experiment})/(\text{standard lifespan})$. The experimental results of Druckrey (8) and Lee and O'Neill (9) suggest that the TD_{50} will be something like f^2d or perhaps even f^3d .

The standard lifespan has been defined to be 24 months for mice and rats. Since most good experiments are scheduled to continue for at least 18 months, f will usually be at least 0.75, and it will therefore not matter greatly whether we correct by f^2 or f^3 . To avoid overcorrection, we recommend that the "corrected" TD_{50} be estimated as f^2d , and we have adopted this convention in the Carcinogenic Potency Database.

Similarly, if an experiment is continued longer than the standard lifespan ($f > 1$), then we recommend that the dose rate d needed to halve the proportion of tumorless animals at the end of the extended period of observation be calculated and the TD_{50} again be estimated as f^2d . (Since few experiments continue past 110 weeks, this correction will not have a large effect.)

Superficially, it would look much more "statistically respectable" to fit a Weibull distribution, in which the incidence rate of tumors was assumed to be proportional to $(\text{duration} - w)^k$, to the experimental results and to estimate the results of lifelong exposure accordingly. Since there are two parameters, a misleadingly excellent fit is assured, but, in fact, degeneracies between the parameters of this formula (and, likewise, between the parameters of various alternative formulae) exist that can cause unpredictable errors (10). This is less true of the suggested routine use of the correction factor f^2 .

For example, let us consider an experiment in which the surviving animals (mice or rats) were sacrificed after 20 months on test. The TD_{50} calculated on the basis of 20 months is then multiplied by the correction factor $f^2 = (20/24)^2 = 0.69$ to yield the TD_{50} based on the "standard" lifespan for the species. Further details of this convention are given in the accompanying paper (4).

Inclusion of Incidental Tumors

When animals die of nonneoplastic disease, or when they are finally sacrificed, some will be found at autopsy to have tumors that may not have been discovered for

several weeks or months if the animals had lived on. In general, actuarial methods should not be applied to "incidental" tumors (3); however, applying these death-rate methods to all tumors will usually lead to only a small error (decrease in estimated probability of remaining tumorless) being made with experiments which end with a terminal sacrifice if all that is wanted is the final proportion of tumorless animals. In practice, these incidental tumors are very important because they may be numerous enough to constitute an appreciable proportion of the information yielded by the whole experiment. Moreover, very few experimental reports, not even the otherwise excellent NCI/NTP records, attempt to distinguish between incidental and nonincidental tumors. The large number of animals suddenly found to have tumors in the final week of an experiment which ends with the sacrifice of all survivors may cause difficulties if Weibull, lognormal, or some other parametric statistical method is used to analyze the data, but need cause no difficulties if nonparametric methods are used.

Selection of Tumor Sites on Which to Base TD_{50}

Consider an experiment comparing a control group and one treated group in which there are two tumor types, only one of which is affected by treatment, and where times to tumor are exponentially distributed. Suppose further that, in the control group, the actuarially adjusted cumulative incidence of tumors at the affected site is 5% and of all tumor-bearing animals is 50%. If the treatment increases the cumulative incidence at the affected site to 15%, this will result in a 55.3% cumulative incidence of all tumor bearing animals. The TD_{50} calculated only for the tumor site affected by treatment will be the same as that calculated for all tumor bearing animals. The advantage of restricting our interest to a single tumor site is that it may yield an analysis which appears more relevant because it is not affected by the random occurrence of unrelated tumors. In addition, there is greater statistical power associated with detecting the 5 to 15% increase than with detecting the 50 to 55.3% increase. In fact, for groups of 50 animals, the expected power is more than four times greater at the lower incidence.

The disadvantage, unless it is absolutely clear which tumors are dose-dependent and which are not, is that such selection automatically biases the comparison of treated with control animals, and thereby tends to exaggerate the carcinogenic potency of the test substance. Moreover, the intuitive reasons for looking only at affected sites are already satisfied by the original definition of the TD_{50} (involving a halving of the number of tumorless survivors), because tumors that have a similar age-specific incidence in all groups of animals, treated or untreated, do not systematically affect the TD_{50} .

If a treatment causes tumors at more than one site,

then the site that is most strongly affected is usually affected much more than the other site(s), and the TD_{50} can be adequately approximated by studying tumors at that single site. If, therefore, a published report describes tumor incidence at certain sites or groups of sites and none of these sites or groups is exactly what is wanted, it is probably best to restrict attention to that reported site or group of sites with the most highly statistically significant carcinogenic effect. As remarked above, in cases of marginal statistical significance, this procedure will exaggerate the carcinogenicity of the test substance, but this will not generally happen when there is a very marked carcinogenic effect. Details of our selection of tissue and tumor types for the Carcinogenic Potency Database are given in the accompanying paper (4).

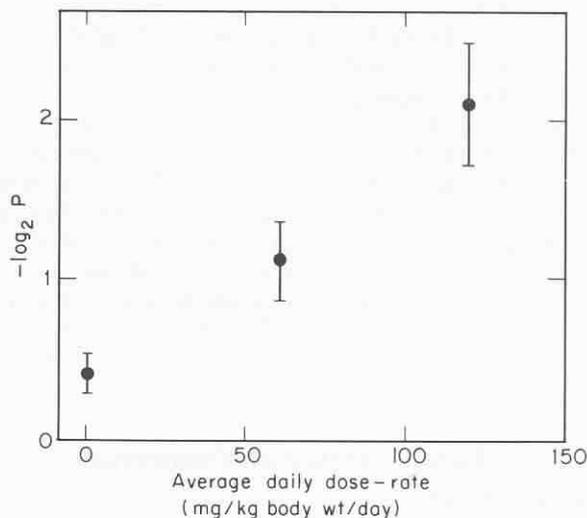
Statistical Methods for Estimating the TD_{50} When Time-to-Tumor Data Are Available

Different statisticians would undoubtedly devise different statistical methods for estimating the TD_{50} . One simple way would be to calculate for each group the probability P of remaining tumorless[‡] and to plot a graph of these probabilities against dose rate. With P plotted on a \log_2 scale, there will then be a unit change in $\log_2 P$ as we go from zero dose to the TD_{50} (Fig. 1).

It may be that such a graph will yield approximately a straight line, for this is predicted by certain rather simple multistage models for cancer induction. However, other equally plausible multistage models do not predict straight lines, and so the expectation that the line might be straight must not distort the interpretation of the actual plotted data.

One simple way to construct a line through such data is to estimate the variance of each $\log_2 P$ value (e.g., by the Greenwood formula (3,11) and then to find which "acceptable" straight line minimizes the inverse-variance-weighted sum of squared deviations of the data from the line. "Acceptable" here means only that the dose-response relationship shall have nonnegative slope (b) and nonnegative intercept (a), so the set of acceptable lines is somewhat constrained. Once a line has been derived, it is easy to read the TD_{50} from it; the slope b equals $1/TD_{50}$, and the confidence interval for the slope yields a confidence interval for $1/TD_{50}$. The only unusual feature about confidence intervals for $1/TD_{50}$ is that if the lower limit is zero or less, then there is no upper confidence limit for the TD_{50} (i.e., we cannot be confident that the substance has any carcinogenic effect). In certain cases, the TD_{50} may represent an impossibly large dose for compounds exhibiting no carcinogenicity; however, we are able to compute a lower confidence limit for such compounds.

[‡]Formally, P is the actuarially adjusted probability of remaining completely free of any fatal or incidental tumor throughout the standard lifespan.



XBL829-4116

FIGURE 1. Graph of $-\log_2 P$ against average daily dose rate, where P = actuarially adjusted probability of remaining tumorless, for NCI/NTP experiment of tris(2,3-dibromopropyl) phosphate in male mice, composite of tissue and tumor pathologies considered positive. Standard errors of $-\log_2 P$ are indicated and were estimated as in Pike and Roe (11).

An alternative to deriving the "best" acceptable straight line by least squares is to derive it by an adaptation (see below) of the methods of Cox (12) for analyzing censored data. Cox's methods are, in various minor ways, preferable for the specific purpose of plotting a straight line through a graph of $-\log_2 P$ and studying its slope, but they are obviously not essential, and for most purposes inverse-variance-weighted least squares lines will be about as good.

Analysis when Time-to-Tumor Data Are Available: Outline of the Statistical Principles Underlying Cox's Methods

Cox (12) pointed out that a likelihood function which allowed standard methods of statistical inference to be used could be derived from a conditional argument: given the numbers of animals at risk of cancer in each group at the start of each week, what is the likelihood of the events that occurred in that week? This conditional argument allows very general parameterization of the time-dependence of risk, usually without material loss of statistical efficiency. We use Cox's arguments to fit the family of models

$$\log_2 P_{ij} = -h_j[a + b(\text{dose}_i)]$$

where a , b , and h_1, \dots, h_t are parameters of the model which are to be fitted subject to $a \geq 0$, $b \geq 0$, and each $h_j \geq 0$; and P_{ij} is the conditional probability of a disease-free animal in group i at time j remaining disease-free until time $j + 1$. Clearly, such a model is degenerate, in that, without changing the predictions for $\log P_{ij}$, the h_j could all be multiplied by any positive constant if a and b were also divided by that constant.

We must therefore normalize the h_j in some way, and one can then usually find unique values \hat{a} , \hat{b} , and \hat{h}_j that maximize Cox's (12) conditional likelihood; \hat{b} then estimates $1/TD_{50}$. Although computationally tedious, the advantages of such methods are that they deal quite naturally with the sudden large numbers of tumors that may be found in the final week of an experiment when all the survivors are killed and autopsied; they are asymptotically efficient against Weibull alternatives; they are robust if Weibull alternatives do not exactly govern the data; and they are rank invariant. Further details of our methods for finding \hat{b} and its confidence limits with this Cox model have been given by Sawyer et al. (13).

Analysis When Actuarial Correction Is Impossible

Often, animal carcinogenicity studies are reported simply in terms of the numbers of tumor-bearing animals in each group; actuarially adjusted numbers are not given, nor are sufficient details of survival and tumor times provided to allow actuarial calculations. As has already been noted, a carcinogenic treatment that also causes deaths by nonneoplastic toxicity may cause so many of the treated animals to die prematurely that the crude proportion of tumor-bearing animals is much smaller than it would have been if actuarial analysis had been possible. This effect is usually most marked among the high-dose animals, and this then results in underestimation of any index of carcinogenicity.

When each data point is simply a binomial proportion, we fit the straight line $-\log_2 P = a + bd$ by a constrained ($a \geq 0$, $b \geq 0$) maximum likelihood fit. Minor difficulties of programming are caused by the possibility that the constrained maximum likelihood value (or one of the confidence limits) may actually lie on one of the constraints $a = 0$ or $b = 0$, but these difficulties can, with due care, be circumvented.

Some authors who do not give full time-to-tumor data nevertheless cite some denominator (e.g., numbers of survivors in each group when the first tumor in the experiment arose) that makes partial allowance for the effects of intercurrent mortality on the numbers at risk of cancer. A crude percentage of tumorless animals based on such a reduced denominator may not be ideal,

but it is often preferable to an even cruder percentage based on the original denominator, as it does at least make some allowance for the effects of premature deaths on the numbers of animals that develop tumors.

This work was supported by DOE Contract DE-AT03-80EV70156 to B.N.A., NIEHS/DOE Interagency Agreement 222-Y01-AS-10066 through the Lawrence Berkeley Laboratory, and Grants (P01 CA 17054) from NCI and (SIG-2) from the American Cancer Society through the University of Southern California School of Medicine.

REFERENCES

1. Finney, D. J. The adjustment for a natural response rate in probit analyses. *Ann. Appl. Biol.* 37: 187-195 (1949).
2. Finney, D. J. *Statistical Method in Biological Assay*. Griffin, London, 1964.
3. Peto, R., Pike, M. C., Day, N. E., Gray, R. G., Lee, P. N., Parish, S., Peto, J., Richards, S., and Wahrendorf, J. Guidelines for simple, sensitive significance tests for carcinogenic effects in long-term animal experiments. In: *IARC Monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Humans. Supplement 2: Long-Term and Short-Term Screening Assays for Carcinogens: A Critical Appraisal*. IARC, Lyon, 1980, pp. 311-426.
4. Gold, L. S., Sawyer, C. B., Magaw, R., Backman, G. M., de Veciana, M., Levinson, R., Hooper, N. K., Havender, W. R., Bernstein, L., Peto, R., Pike, M. C., and Ames, B. N. A carcinogenic potency database of the standardized results of animal bioassays. *Environ. Health Perspect.* 58: 9-319 (1984).
5. Meselson, M., and Russell, K. Comparisons of carcinogenic and mutagenic potency. In: *Origins of Human Cancer, Book C* (H. H. Hiatt, J. D. Watson and J. A. Winsten, Eds.), Cold Spring Harbor Laboratories, Cold Spring Harbor, NY, 1977, pp. 1473-1481.
6. Crouch, E., and Wilson, R. Interspecies comparison of carcinogenic potency. *J. Toxicol. Environ. Health* 5: 1095-1118 (1979).
7. National Academy of Sciences. *Regulating Pesticides*. Committee on Prototype Explicit Analyses for Pesticides, National Research Council, Washington, DC, 1980.
8. Druckrey, H. *Quantitative Aspects in Chemical Carcinogenesis* (UICC Monograph Series, Vol. 7). Springer-Verlag, Berlin, 1967, pp. 60-77.
9. Lee, P. N., and O'Neill, J. A. The effect both of time and of dose applied on tumor incidence rate in benzpyrene skin painting experiments. *Brit. J. Cancer* 25: 759-770 (1971).
10. Peto, R., and Lee, P. N. Weibull distributions for continuous-carcinogenesis experiments. *Biometrics* 29: 457-470 (1973).
11. Pike, M. C., and Roe, F. J. C. An actuarial method of analysis of an experiment in two-stage carcinogenesis. *Brit. J. Cancer* 17: 605-610 (1963).
12. Cox, D. R. Regression models and life tables. *J. Roy. Statist. Soc. B32*: 187-220 (1972).
13. Sawyer, C., Peto, R., Bernstein, L., and Pike, M. C. Calculation of carcinogenic potency from long-term animal carcinogenesis experiments. *Biometrics* 40: 27-40 (1984).