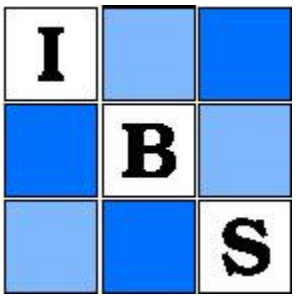

BIOMETRICS 49, 917–926
September 1993

A Measure of Tumorigenic Potency Incorporating Dose–Response Shape

Kristen L. Meier,[1,*] A. John Bailer,[2,3] and Christopher J. Portier[1]

[1] Risk Methodology Section, Statistics and Biomathematics Branch, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina 27709, U.S.A.
[2] Department of Mathematics and Statistics, Miami University, Oxford, Ohio 45056, U.S.A.
[3] Division of Standards Development and Technology Transfer, National Institute for Occupational Safety and Health, 4676 Columbia Parkway, Cincinnati, Ohio 45266, U.S.A.

SUMMARY

Many researchers have considered the problem of ranking chemical agents based on their carcinogenic potency. Sawyer et al. (1984, *Biometrics* **40**, 27–40) proposed a carcinogenic potency estimate that incorporates both intercurrent mortality and background tumor rates. Since then, many authors have either generalized the method outlined by Sawyer et al. or developed their own method based on a slightly different adjustment for treatment-related changes in survival. None of these methods, however, has estimated the shape of the dose–response curve and incorporated such an estimate in potency estimation. In this manuscript, a measure of tumorigenic potency is proposed that utilizes the estimated shape of the dose–response relationship, in addition to estimated dose effects, in order to rank chemicals on the basis of carcinogenic risk. Comparison of this new measure to that of Sawyer et al. is done using a large database of animal carcinogenicity experiments from the National Cancer Institute and the National Toxicology Program.

1. Introduction

Many experiments have demonstrated that chemical agents vary considerably in the dose required to induce tumors in laboratory animals. It has been shown that the same holds true in humans (Allen, Crump, and Shipp, 1988). This suggests that certain substances are more "potent" carcinogens than others; i.e., they are able to induce tumor formation at lower doses than others. One would like to measure the potency of a substance and to rank several substances based on this potency measure so that results from different experiments can be compared quantitatively and substances can be ordered according to the potential hazards involved in their use (OSHA, 1983).

The "tumorigenic dose 50" (TD_{50}), analogous to the "lethal dose 50" (LD_{50}), was first introduced by Sawyer et al. (1984) and utilized by Peto et al. (1984) and Gold et al. (1984) as a measure of tumorigenic potency of a chemical in long-term animal carcinogenicity experiments. The TD_{50} is defined by Sawyer et al. (1984, p. 27) as the "daily dose rate required to halve the probability of remaining tumorless at the end of a standard life-span." This potency measure assumes that the hazard of tumor onset is linearly related to dose (a linear proportional hazard assumption) and that tumor onset times can be observed.

The data from which potency measures have traditionally been estimated consist of the percentage of tumor-bearing animals at different dose groups from a chronic bioassay and it may consist of additional information such as cause of death or actual time of tumor onset. Sawyer et al. (1984) were the first to propose a potency measure, the TD_{50}, where the percentages in each dose group were adjusted for both the incidence of spontaneous tumors and intercurrent mortality. The importance

* *Current address:* Statistics and Data Systems Division, Food Safety and Inspection Service/USDA, Washington, D.C. 20250, U.S.A.

Key words: Carcinogenicity studies; Carcinogenic potency; Intercurrent mortality; Risk assessment; Survival analysis; TD50; Tumor incidence.

of these survival adjustments when analyzing carcinogenicity experiments has been identified by various researchers including Hoel and Walburg (1972). Intercurrent mortality includes premature deaths due to causes unrelated to the tumor of interest, for example, deaths due to the toxicity of the administered compound. The latter is particularly prevalent in the high-dose groups where percentages that are not adjusted for competing causes of death may be artificially low when the dosed animal dies before it has time to develop a tumor.

Another factor that complicates the estimation of the TD_{50} is tumor lethality. If the time of onset of a tumor is unobservable for an animal dying during the course of the study (an occurrence that is quite common), then in order for the TD_{50} to estimate tumor incidence, one must assume, under Sawyer's model, that the tumor is instantly lethal. This implies that the time of death with tumor present equals the time of tumor onset. If this is not the case and the TD_{50} is still used as a potency measure, then this measure of tumor incidence is confounded with mortality and biased potency estimates are likely to result (Portier and Hoel, 1987).

In many carcinogenicity experiments, the presence of a tumor does not significantly alter survival of the experimental animal (Portier, Hedges, and Hoel, 1986). In such cases, the TD_{50} is related to the rate of death with tumor rather than the tumor incidence rate. This is in direct conflict with the goal of the carcinogenicity study, which is to evaluate tumor incidence (McKnight and Crowley, 1984). Portier and Hoel (1987) have shown that the TD_{50} is extremely sensitive to the tumor lethality assumption and suggest that "potency estimation could be improved by basing the [TD_{50}] estimates upon the tumor incidence rate rather than upon the tumor death rate when an estimate of tumor lethality is obtainable."

As a result, various attempts have been made to derive a potency measure that is not dependent on the tumor lethality assumption. Finkelstein (1991) and Finkelstein and Ryan (1987) propose alternative methods that do not rely on tumor lethality assumptions, but do assume that the cause of death for each animal is known. Bailer and Portier (1993) propose a carcinogenic potency measure that applies to tumors of all lethality types and that assumes cause of death information is unavailable or, when available, is not necessary. This measure avoids confounding of the tumor incidence rate with intercurrent mortality by assuming that the shape and form (Weibull) of the tumor onset distribution are known. Like Sawyer et al. (1984), Bailer and Portier (1993) use the linear proportional hazards assumption.

In each of the works cited, the research efforts have focused on adjusting the response in each dose group rather than on examining the reasonableness of the model imposed on the dose–response relationship. There have been some recent investigations into estimating the shape of dose–response curves in carcinogenicity studies (Bailar et al., 1988; Williams and Portier, unpublished manuscript; Hoel and Portier, 1993). This work has focused on the range and frequency of different shapes, however, not on potency estimation.

In this paper, a measure of carcinogenic potency is considered that attempts to incorporate an estimate of the shape of the dose–response curve. Since most human exposures to chemical agents of environmental concern are at low doses, the shape of the dose–response curve is very important. We propose a potency measure that explicitly incorporates an estimate of this shape. This new measure is compared to currently applied potency measures. The comparison is done using data from numerous carcinogenicity studies and through analytical derivations.

2. Notation and Methodology

2.1 *The Bioassay*

The database considered for analysis is a subset of data collected through the Carcinogenesis Bioassay Program of the National Cancer Institute/National Toxicology Program (NCI/NTP). In the "standard NCI bioassay," both sexes of rats and mice are exposed to some compound for most of their lifetime (usually 2 years). The doses administered in such an experiment are determined in a smaller (3- to 6-month) study. From this smaller study, the "maximum tolerated dose" (MTD) of the compound for each sex–species group is determined. NCI defines the MTD as the maximum level of exposure "that can be predicted not to alter the animals' normal longevity from effects other than carcinogenicity" (Sontag, Page, and Saffiotti, 1975). In a typical bioassay, groups of 50 animals each are exposed, for 104 weeks, to one of several doses of the compound, including a zero dose (control), and the MTD. For the data we examined, the number of experimental groups/dose levels ranged from 3 to 7. We denote the dose levels examined as $d_0 = 0$ (control), $d_1, d_2, \ldots, d_m$ with n_i animals exposed to dose d_i ($i = 0, \ldots, m$). For the jth animal in the ith dose group ($j = 1, \ldots, n_i$; $i = 0, \ldots, m$), δ_{ij}, the presence ($\delta_{ij} = 1$) or absence ($\delta_{ij} = 0$) of a tumor, along with t_{ij}, the survival time, was recorded.

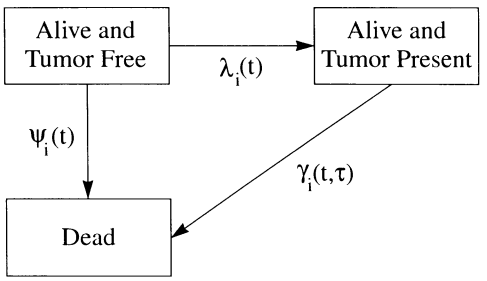

Figure 1. A three-state model for a tumorigenicity experiment.

2.2 *The Mathematical Model*

The proposed potency estimator is an extension of the estimator in Bailer and Portier (1993). These authors use a three-state compartment model (see Figure 1) for tumor onset in which an animal starts tumor-free, but incurs a hazard over time of tumor onset, $\lambda(t)$, and a competing hazard of death without tumor, $\psi(t)$. If tumor onset occurs at time τ, then the transition from the tumor-bearing live state to the death state is governed by the hazard, $\gamma(t, \tau)$. Let E_1 be the random variable that represents the time from initial exposure to the occurrence of the first event (either tumor onset or tumor-free death), and let E_2 represent the time until death with tumor. Further, let δ be an indicator of tumor presence ($\delta = 1$) or absence ($\delta = 0$). The hazard functions can be explicitly defined as

$$\lambda_i(t) = \lim_{\Delta \to 0} \Delta^{-1}\Pr\{t \leqslant E_1 < t + \Delta,\ \delta = 1 \mid E_1 \geqslant t,\ d_i\},$$

$$\psi_i(t) = \lim_{\Delta \to 0} \Delta^{-1}\Pr\{t \leqslant E_1 < t + \Delta,\ \delta = 0 \mid E_1 \geqslant t,\ d_i\},$$

$$\gamma_i(t, \tau) = \lim_{\Delta \to 0} \Delta^{-1}\Pr\{t \leqslant E_2 < t + \Delta \mid E_1 = \tau \leqslant t,\ \delta = 1,\ E_2 \geqslant t,\ d_i\},$$

where i denotes the dose group, $i = 0, 1, 2, \ldots, m$.

For a fixed standard lifespan or terminal sacrifice time, TS, if we denote $q_i = \Pr$(animal at dose d_i has not been diagnosed with a tumor by time TS, in the absence of all other causes of death), then from survival analysis,

$$q_i = \exp\left(-\int_0^{TS} \lambda_i(t)\, dt\right).$$

We adopt the conventional "standard lifespan" for rats and mice, which is TS = 2 years (104 weeks) (Peto et al., 1984). A common assumption is that tumor incidence can be factored into a linear function of dose and a function of age, i.e.,

$$\lambda_i(t) = (1 + \beta d_i)\lambda_0(t),$$

where $\lambda_0(t)$ is the tumor incidence rate at age t for animals in the control group (Sawyer et al., 1984). We call this the linear proportional hazard assumption. An immediate result of this assumption is that q_i can be reexpressed as

$$q_i = (q_0)^{1+\beta d_i}.$$

As stated earlier, much research has focused on how to best estimate the q_i's, and little attention has been given to studying the validity of the linear proportional hazard assumption.

To better understand the implications of this assumption, we reparameterize the model for q_i (Bailer and Portier, 1993). Letting $q_0 = \exp(-\alpha)$ and $\beta = \beta^*/\alpha$, we can write

$$q_i = \exp(-\alpha - \beta^* d_i),$$

where $\exp(-\alpha)$ corresponds to the background probability of remaining tumor-free in the absence of competing risks and $\exp(\beta^*)$ is the factor by which this background probability is divided with each unit increase in dose. A common way of displaying a dose–response relationship graphically is in terms of $p_i = 1 - q_i$ versus dose, d_i. The dose–response curve is then given by $p_i = 1 - \exp(-\alpha - \beta^* d_i)$, a simple linear, complementary log model. The shape of this dose–response curve is a monotone increasing, concave downward function (i.e., with a monotone decreasing derivative). A

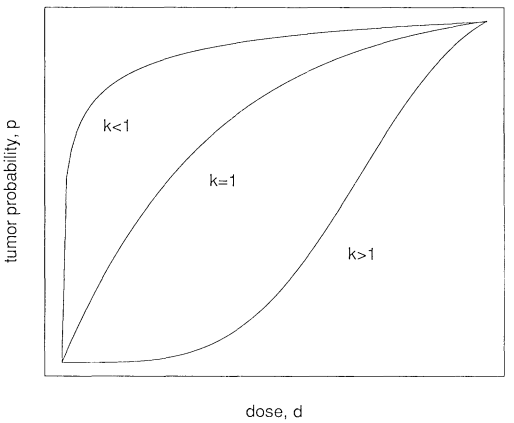

Figure 2. Possible shapes of a dose–response curve generated from the model $p_i = 1 - \exp(-\alpha - \beta^* d_i^k)$.

graph of such a curve is given in Figure 2 (see curve labelled $k = 1$). Upon examination of the NCI/NTP data, however, it appears that such a shape is inadequate to describe the range of possible dose–response relationships (Hoel and Portier, 1993). The driving force behind this particular shape is the linear proportional hazards assumption. If the relationship between tumor incidence and dose was quadratic, cubic, or some higher-degree polynomial, then the dose–response curve could describe a wider range of relationships.

We propose the following relationship between tumor incidence and dose:

$$q_i = \exp(-\alpha - \beta^* d_i^k), \tag{1}$$

where $\alpha \geq 0$, $\beta^* \geq 0$, and $0 \leq k \leq 10$. These constraints reflect the fact that we are focusing our interest on toxic agents rather than on therapeutic agents. We chose the maximum value of $k = 10$ for computational convenience. Qualitatively, models where $k > 10$ are indistinguishable from models with $k = 10$ so little information on shape is lost by restricting our analysis to $k \leq 10$. Relationship (1) is based on the assumption that

$$\lambda_i(t) = (1 + \beta d_i^k)\lambda_0(t). \tag{2}$$

Such an assumption allows for a more flexible dose–response relationship, yet it introduces only one new parameter, k. This parameter will be referred to as the shape parameter. We will refer to model (1) as the power proportional hazard model since dose is raised to the kth power. Note that under this model, when $k > 1$ the dose–response curve is concave upward or sublinear (threshold-like) at low doses, and when $k < 1$, the curve is concave down or supralinear (see Figure 2). Assuming that the model is correct, in the latter case carcinogenic risk at low doses would be underestimated under the incorrect linear model and in the former case carcinogenic risk at low doses would be overestimated.

The goal is to estimate (α, β^*, k) in (1). We use a likelihood assuming binomial sampling of a survival-adjusted population. The likelihood can be written as

$$L \propto \prod_{i=0}^{m} (p_i)^{x_i}(q_i)^{n_i^* - x_i},$$

where x_i is the number of animals in the ith dose group with the tumor present, n_i is the number of animals in the ith dose group, and n_i^* is the survival-adjusted number of animals in the ith dose group. The survival-adjusted number of animals in the group is calculated in the manner of Bailer and Portier (1988, 1993). Briefly, if δ_{ij} is an indicator variable of tumor presence (= 1 if tumor present, 0 if not) for the jth animal in the ith dose group, then $n_i^* = \sum_j \omega_{ij}(1 - \delta_{ij}) + \sum_j \delta_{ij}$, where j varies from 1 to n_i and $\omega_{ij} = (t_{ij}/\mathrm{TS})^3$. In essence, this method assigns partial information to animals that die prior to sacrifice and fail to get the tumor prior to death. Under the assumption that the rate of tumor incidence is proportional to t^2, Bailer and Porter (1988) have shown that this adjustment approximates p_i. The likelihood is then based on the assumption that $x_i \sim \mathrm{bin}(n_i^*, p_i)$. For additional justification of the use of the correction term ω_{ij}, see Bailer and Portier (1988) and Portier et al. (1986).

Substituting equation (1) into the likelihood above, and then taking the natural logarithm, the log-likelihood, less a constant equals

$$\ln(L) = \sum_{i=0}^{m} \{x_i \ln[1 - \exp(-\alpha - \beta^* d_i^k] - (n_i^* - x_i)(\alpha + \beta^* d_i^k)\}. \tag{3}$$

An optimization algorithm (Davidson–Fletcher–Powell method; see Walsh, 1975, Chap. 5) is used to obtain the maximum likelihood estimates (MLEs) for α, β^*, and k. We refer to these estimates and potency measures computed using these estimates as being TI-based estimates since they are based on tumor incidence rather than tumor death. Those estimates derived using the methods of Sawyer et al. (1984) will be called TD-based estimates.

2.3 *Estimating the Tumorigenic Potency*

One potency measure that can be derived is the TI_ρ, which is analogous to the TD_ρ, a generalization of the TD_{50} (Sawyer et al., 1984). The TI_ρ is that dose for which the probability of remaining tumor-free, in the absence of competing risks, is ρ times q_0. That is, it is the dose D such that $\rho q_0 = (q_0)^{1+\beta D^K}$. Solving for D yields

$$D = \{\ln(\rho)/[\beta \ln(q_0)]\}^{1/k},$$

or simply

$$D = \{\ln(\rho)/-\beta^*\}^{1/k} \tag{4}$$

using the parameterization $q_0 = \exp(-\alpha)$ and $\beta = \beta^*/\alpha$. Note that the first form for D is analogous to Sawyer's equation (6) when $k = 1$. An estimate of the TI_ρ is obtained by substituting the MLEs for k and β^* into the right-hand side of (4). The differences between our estimate of the TI_ρ and Sawyer's estimate of the TD_ρ are that our method allows for a more general dose–response relationship and that our estimate depends only on tumor incidence.

Under the linear proportional hazards assumption, i.e., when $k = 1$, the relative ranking of a set of compounds based on the TI_ρ is independent of the choice for ρ. This follows from the fact that

$$TI_{\rho_1} = \{\ln(\rho_1)/\ln(\rho_2)\}TI_{\rho_2},$$

where $\ln(\rho_1)/\ln(\rho_2)$ does not depend on a particular compound. This is equally true for the TD_ρ. Under the more general power model where k is allowed to vary, however, the choice of ρ becomes crucial. One can write

$$TI_{\rho_1} = \{\ln(\rho_1)/\ln(\rho_2)\}^{1/k}TI_{\rho_2},$$

where the multiplier, $\{\ln(\rho_1)/\ln(\rho_2)\}^{1/k}$, now depends on the shape parameter corresponding to a particular compound. This issue is discussed in more detail in Section 3.2 below.

3. Application to NCI/NTP Bioassay Data

We analyzed 1,577 different data sets on 286 chemicals studied in the Carcinogenesis Bioassay Program of the National Cancer Institute/National Toxicology Program (NCI/NTP). The 1,577 data sets were chosen because they demonstrated a statistically significant ($\alpha = .05$) dose–response relationship as determined from the one-sided trend test of Bailer and Portier (1988). The 1,577 data sets contain information on 263 different tumor sites, on primarily two different species of both male and female rodents (F344 rats and $B6C3F_1$ mice).

3.1 *Power Model Versus Linear Model*

We fit both the power proportional hazards model (k estimated) and the linear proportional hazards model ($k = 1$) to each of the 1,577 data sets using the methods described in Section 2.2. There were 442 (28%) data sets out of 1,577 with estimates of k less than 1, and 1,135 (72%) with estimates of k greater than 1; k was estimated as 0 (i.e., on the lower boundary of the parameter space) in 107 (7%) of the 1,577 data sets.

When k equals zero, the estimated dose–response curve is a step function with (background) response $1 - \exp(-\alpha)$ at dose = 0, and $1 - \exp(-\alpha - \beta)$ at doses greater than zero. As a result, the TI_ρ measure is estimated to be zero. We examined these 107 data sets and found that in each case, there was an initial increase in dose–response and then a (usually small) decrease. Overall, the trend test was still significant, but the downturn in dose–response could not really be captured by the model; the best model that could incorporate this decrease was one where k is zero. The linear model

Table 1
A summary of the distribution of the estimated shape parameter, k

	$0 < k < 1$	$1 < k < 10$	$k = 10$	Totals[a]
All data sets	335 (22.8%)	594 (40.4%)	541 (36.8%)	1,470
Reject $k = 1$[b]	9 (5.5%)	76 (46.3%)	79 (48.2%)	164

[a] Data sets where k is estimated as zero have been omitted.
[b] This row pertains to those data sets where the shape parameter is significantly different from 1.

could still be fit to such data, but the fit appeared inadequate visually. Under both the linear and the power model, it is assumed that the true risk is increasing in dose. Failure to account for a downturn in the data is a shortcoming of all of the methods discussed, including our proposed method. A downturn in response may be due to the fact that the MTD used in the experiment was too high or that the toxicity of the compound resulted in internal damage, but not death, thereby affecting tumor growth (Brown, 1985). A new experiment, with different (smaller) dose levels might be appropriate. We believe that the power model is adequate for the remaining 1,470 data sets, and that it is more appropriate than the linear model. Data sets yielding estimates of $k = 0$ have been excluded from the analysis, leaving 1,470 data sets for analysis.

The difference between the power proportional hazards model and the linear proportional hazards model was examined by performing a likelihood ratio test. Specifically, we tested the null hypothesis that $k = 1$ under the assumption of model (1) versus the alternative, $k \neq 1$. We computed the test statistic $2(\ln L_1 - \ln L_0)$, where L_i is the likelihood described in (3) under the null ($i = 0$) or alternative ($i = 1$) hypothesis, and compared it to the 95th percentile of a chi-square distribution with 1 degree of freedom (Cox and Hinkley, 1974). The null hypothesis was rejected for 164 of the 1,470 tests or approximately 11% of the time (if $\alpha = .10$, we reject in 302 cases or 20% of the time). Nine (5%) of the 164 rejections had estimates of k less than 1, and 155 (95%) had estimates greater than 1. A more detailed summary of these results is given in Table 1. The median value of the estimated k's for the 164 data sets where the linear model was rejected is 5.77; the median of the k's for all 1,470 data sets is 2.36. These results indicate that allowing for a convex curve is appropriate.

3.2 *Behavior of Potency Measures*

The choice of ρ in TI_ρ is an important consideration when doing potency estimation. The TI_{01} appears to be a particularly reasonable measure of tumorigenic potency for the data we considered here for two reasons. First, in risk assessment, regulators are interested in very small increases to lifetime risk. They are interested in the effects of the compound at low doses. A second reason, discussed in more detail below, is that unlike the commonly used TD_{50} or TI_{50}, the TI_{01} is more likely to lie below the MTD. In the spirit of not extrapolating beyond the range of the data, it seems reasonable that a desirable potency estimate lies below the maximum dose tested. Perhaps most important, we wish to emphasize that the measure should take into account the shape of the dose–response curve.

Potency estimates are frequently used to determine priorities for regulating carcinogens. In this application, the rank order of potency estimates is more useful than the actual potency estimates. A debate in risk analysis literature has arisen from basing such rankings on the TD_{50} (Ames, Magaw, and Gold, 1987; Wartenberg and Gallo, 1990). We now explore the differences in rankings that might occur when using TD-based potency estimators versus TI-based potency estimators. In Section 2.3 it was pointed out that when using the TI_ρ as a measure to rank tumorigenic potencies of compounds, different choices of ρ will yield different rankings under the power model. (Recall that this is not the case under the linear model.) This led us to examine several tumorigenic potency estimates for each of the 1,470 ($k = 0$ omitted) data sets examined. Definitions of two TI_ρ measures ($\rho = .01, .50$) and the TD_{50}, the primary measures we considered, are given in Table 2. TI_ρ measures for $\rho = .10, .05, .001$, and .000001 were also examined. All TI_ρ measures were computed based on the Poly-3 adjusted tumor counts (Bailer and Portier, 1988), and all were computed at (or adjusted to) 104 weeks. In addition, each TI measure was computed twice, first under the assumption of linear proportional hazards and then under the assumption of power proportional hazards. To distinguish between the two, we adopt the notation $TI_{01,lin}$ to represent the TI_{01} computed under the assumption of linear proportional hazards, and $TI_{01,pow}$ to represent the TI_{01} computed under the assumption of power proportional hazards, etc.

There were 282 data sets for which we were unable to use the TD_{50} potency measure. These cases arose primarily for data sets with staggered terminal sacrifice times (i.e., different treatment groups

Table 2
Definition of potency measures computed

TI_{01}	Dose required to decrease the probability of remaining tumor-free, in the absence of competing risks, by 1%; computed using the methods described in Section 2.2
TI_{50}	Dose required to decrease the probability of remaining tumor-free, in the absence of competing risks, by 50%; computed using the methods described in Section 2.2
TD_{50}	Dose required to decrease the probability of remaining tumor-free, in the absence of competing risks, by 50%; computed using the methods of Sawyer et al. (1984)

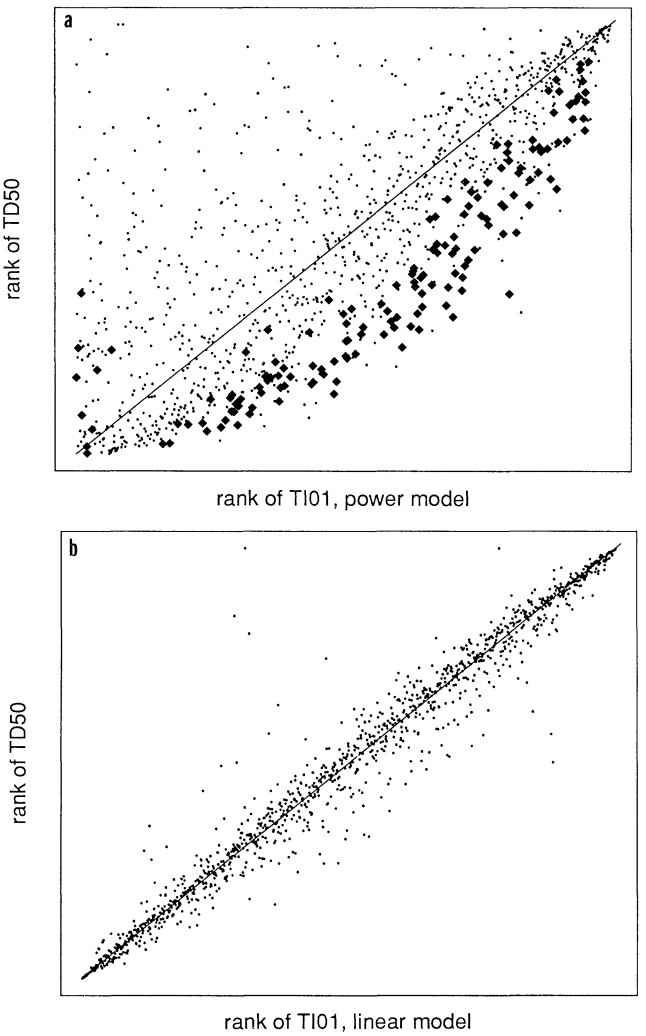

Figure 3. (**a**) A plot of the rank of the TD_{50} versus the rank of the $TI_{01,pow}$ computed for each of 1,188 data sets. The reference line results if the two measures yield the same rank. Points highlighted by solid diamonds correspond to data sets where the shape parameter, k, is significantly different from 1. (**b**) A plot of the rank of the TD_{50} versus the rank of the $TI_{01,lin}$ computed for each of 1,188 data sets.

were sacrificed at different times) and with clustering of tumor responses at these times. Any analysis including the TD_{50} utilizes only the 1,188 data sets for which the TD_{50} estimate was finite.

Figures 3a and 3b display a plot of the ranks of the TD_{50} versus the corresponding ranks of the $TI_{01,pow}$ and a plot of the ranks of the TD_{50} versus the corresponding ranks of the $TI_{01,lin}$ for 1,188 data sets. The points in Figure 3a are much more scattered than those in Figure 3b, indicating that the addition of the power parameter is associated with the greater change in the ranks of the TI_{01} measures versus the TD_{50}. We interpret the differences of scatter in these two figures as being due to

the poor fit of the linear model for many of the data sets. In Figure 3a, the 199 points corresponding to data sets where the shape parameter, k, is significantly different from 1 are marked by solid diamonds. Points on this plot below the reference line, near the edge of the point cluster, correspond to data sets with large estimates of k ($k > 1$); those far above the line correspond to data sets with small estimates of k ($k < 1$). In this figure, note that low ranking (more potent) compounds based on the $TI_{01,pow}$ do not necessarily have a correspondingly low rank based on the TD_{50}, even when the shape parameter is not significantly different from 1. Compounds ranked high (less potent) using the $TI_{01,pow}$, on the other hand, tend to be ranked high using the TD_{50} as a potency measure. The choice of the measure will determine whether a compound is considered potent.

As stated earlier, one justification for the use of the $TI_{01,pow}$ over other measures (e.g., $\rho < .01$) is that unlike other measures, it is always less than the MTD. For each measure listed in Table 3, we computed the number of times out of the 1,470 data sets (1,188 for the TD_{50}) that the estimated potency measure was outside the range of the data (i.e., the number of times that the measure was greater than the MTD). Note that only the TI_{01} potency measures always lie below the MTD.

Another advantage of using the $TI_{01,pow}$ over other potency measures is that, compared to the other measures, it has the smallest correlation with the MTD. Bernstein et al. (1985) first noted that in the data they examined, the TD_{50} and the MTD were positively correlated. We also found this to be true in the data we examined. Table 4 illustrates the point from Kodell, Gaylor, and Chen (1991), that the shape of the dose–response curve is more important than the experimental design considerations (e.g., the value of the MTD) for determining the behavior of the potency estimate. In Table 4, a summary of different measures of association between various tumorigenic potency estimators and the MTD is provided. Correlations were computed using the logs of the measures because the difference between the minimum and maximum measure is at least five (not including outliers) orders of magnitude. We computed Pearson's correlation coefficient twice, once using all of the data, and once using a portion of the data. The points omitted were high leverage points corresponding to the compound, dioxin, and outliers. Points were considered outliers if they were greater than 20 orders of magnitude from the main cluster of points. These points correspond to cases where k is estimated to very close to zero. In these data sets, if the maximum observed risk is small (slightly greater than or less than ρ), then the estimated potency measure is very large. Alternatively, if the maximum observed risk is large then the estimated potency measure is near zero. The points

Table 3
How often various potency measures lie above the highest dose tested (MTD)

Potency measure	Number > MTD	Percent[a]
TD_{50}	893	75.2%
$TI_{50,pow}$	1,141	77.6%
$TI_{10,pow}$	495	33.7%
$TI_{05,pow}$	179	12.2%
$TI_{01,pow}$	0	0%
$TI_{50,lin}$	1,225	83.3%
$TI_{01,lin}$	0	0%

[a] Percents for TI measures are based on 1,470 data sets; the TD_{50} is based on 1,188 data sets.

Table 4
Associations between the $\log_{10}$(MTD) *and the* $\log_{10}$ *of various potency measures*

$\log_{10}$(MTD) versus $\log_{10}$:	Pearson's correlation (all data[a])	Pearson's correlation (points removed[b])	Spearman's rank correlation (all data)
TD_{50}	.786	.841	.828
$TI_{50,lin}$	.888	.843	.834
$TI_{01,lin}$			
$TI_{50,pow}$	.440	.686	.842
$TI_{01,pow}$	.237	.394	.662

[a] The correlation for the TD_{50} is based on 1,188 points; all others are based on 1,470 points.

[b] Outliers and points corresponding to dioxin were removed. The numbers of outliers removed from each of the four categories in the table, from top to bottom, are 2, 0, 5, and 10. In addition, 23 high-leverage points corresponding to dioxin were removed from each category.

corresponding to dioxin were removed because the corresponding MTD is at least four orders of magnitude greater than the MTD for all other compounds and we did not want the measure of association to be driven by this one compound. We computed Spearman's rank correlation to provide a robust measure of association.

The greatest differences between Pearson's correlation and Spearman's rank correlation occur with the $TI_{\rho,pow}$-based measures. Given the results in the first column of Table 4, one might think that the correlation with the MTD is greatly reduced by using a TI_{50} assuming a power model (correlation = .440) rather than a linear model (correlation = .888). Upon examination of the Spearman rank correlations, however, this phenomenon disappears. The greatest difference in the correlations between the MTD and the $TI_{\rho,lin}$ versus the MTD and $TI_{\rho,pow}$ are when $\rho = .01$. Regardless of which correlation measure is used, the MTD is correlated least with the $TI_{01,pow}$.

4. Concluding Remarks

The use of a power proportional hazards assumption allows for a wider variety of dose–response relationships than does the use of a linear proportional hazard assumption. The former allows for increasing, sigmoidal curves whereas the latter is restricted to a family of curves that are increasing and concave. When estimating the potency association with a low level of risk, the type of potency measure used can have a great effect.

For reasons already discussed, there are many benefits to using a $TI_{01,pow}$ over other measures as a measure of tumorigenic potency. In Section 3.3 it was shown that while the $TI_{01,pow}$ has the smallest correlation with the MTD out of those measures considered, it is still nontrivially correlated with the MTD. It may be appropriate to compare potencies only for those compounds that have similar MTDs, i.e., stratify on some measure of dose/toxicity (cf. Piegorsch et al., 1992). While the MTD does not give direct information regarding the tumorigenicity of a compound, it does give information about the toxicity of a compound.

We have examined other $TI_{\rho,pow}$ measures for choices of $\rho < .01$. It is not clear, however, which measure is "best" without knowing the exact context in which the measure is used. In the brief analysis that we have done, we believe that a $TI_{01,pow}$ measure is an appropriate potency measure, and is desirable over a TD_{50} or $TI_{01,lin}$.

Finally, it is important to keep in mind that potency measures such as those discussed in this paper give just one indication about the potency of a compound. Ideally, a regulatory decision regarding the potency of a compound would incorporate information on mechanisms, toxicokinetics, and epidemiological studies.

Acknowledgements

We thank two anonymous reviewers for helpful comments. A. J. Bailer's work was partially supported by an Ohio Board of Reagents Challenge Grant and a Miami University College of Arts and Sciences Small Research Grant.

Résumé

Beaucoup de chercheurs ont considéré le problème de classer des agents chimiques en fonction de leur pouvoir cancérigène. Sawyer et al. (1984, *Biometrics* **40,** 27–40) a proposé une estimation du pouvoir cancérigène incorporant à la fois la mortalité intercurrente et l'incidence des tumeurs. Depuis, plusieurs chercheurs ont soit généralisé la méthode proposée par Sawyer et al. ou développé leur propre méthode à partir d'un ajustement pour les modifications de la survie liées au traitement légèrement différent. Aucune de ces méthodes, cependant, n'a estimé la forme de la courbe dose–response et n'a incorporé cette estimation dans l'évaluation du pouvoir cancérigène. Dans cet article, une mesure du pouvoir cancérigène est proposée qui utilise la forme estimée de la courbe dose–response en plus de l'effet estimé de chaque dose, afin de classer les agents chimiques en fonction du risque carcérigène. Une comparaison de cette nouvelle mesure avec celle de Sawyer et al. est faite en utilisant une base de données des essais de cancérogénécité du National Cancer Institute et du National Toxicology Program.

References

Allen, B. C., Crump, K. S., and Shipp, A. M. (1988). Correlation between carcinogenic potency of chemicals in animals and humans. *Risk Analysis* **8,** 531–544.

Ames, B. L., Magaw, R., and Gold, L. S. (1987). Ranking possible carcinogenic hazards. *Science* **236,** 271–280.

Bailar, J. C. III, Crouch, E. A. C., Shaikh, R., and Spiegelman, D. (1988). One-hit models of carcinogenesis: Conservative or not? *Risk Analysis* **8,** 485–497.

Bailer, A. J. and Portier, C. J. (1988). Effects of treatment-induced mortality and tumor-induced mortality on tests for carcinogenicity in small samples. *Biometrics* **44,** 417–432.

Bailer, A. J. and Portier, C. J. (1993). An index of tumorigenic potency. *Biometrics* **49,** 357–365.

Bernstein, L., Gold, L., Ames, B., Pike, M., and Hoel, D. (1985). Some tautologous aspects of the comparison of carcinogenic potency in rats and mice. *Fundamental and Applied Toxicology* **5,** 79–86.

Brown, S. L. (1985). Quantitative risk assessment of environmental hazards. *Annual Review of Public Health* **6,** 247–267.

Cox, D. R. and Hinkley, D. V. (1974). *Theoretical Statistics.* London: Chapman and Hall.

Finkelstein, D. M. (1991). Modeling the effect of dose on the lifetime tumor rate from an animal carcinogenicity experiment. *Biometrics* **47,** 669–680.

Finkelstein, D. M. and Ryan, L. M. (1987). Estimating carcinogenic potency from a rodent tumorigenicity experiment. *Journal of the Royal Statistical Society, Series C* **36,** 121–133.

Gold, L. S., Sawyer, C. B., Magaw, R., Buckman, G. M., deVecidna, M., Levinson, R., Hooper, N. K., Havender, W. R., Bernstein, L., Peto, R., Pike, M. C., and Ames, B. N. (1984). A carcinogenic potency database of the standardized results of animal bioassays. *Environmental Health Perspectives* **58,** 9–319.

Hoel, D. and Portier, C. (1993). Nonlinearity of carcinogenesis dose–response functions. *Environmental Health Perspectives* to appear.

Hoel, D. and Walburg, H. (1972). Statistical analysis of survival experiments. *Journal of the National Cancer Institute* **49,** 362–372.

Kodell, R. L., Gaylor, D. W., and Chen, J. J. (1991). Carcinogenic potency correlations: Real or artifactual? *Journal of Toxicology and Environmental Health* **32,** 1–9.

McKnight, B. and Crowley, J. (1984). Tests for differences in tumor incidence based on animal carcinogenesis experiments. *Journal of the American Statistical Association* **79,** 639–648.

National Cancer Institute and National Toxicology Program Bioassay Program. Technical Report Nos. 1–355. Washington, D.C.: U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health.

OSHA (1983). *Hazard Communication: Final Rule.* Department of Labor, Fed Reg 48: 53280-53348.

Peto, R., Pike, M. C., Bernstein, L., Gold, L. S., and Ames, B. N. (1984). The TD_{50}: A proposed general convention for the numerical description of the carcinogenic potency of chemicals in chronic-exposure animal experiments. *Environmental Health Perspectives* **58,** 1–8.

Piegorsch, W. W., Carr, G. J., Portier, C. J., and Hoel, D. G. (1992). Concordance of carcinogenic response between rodent species: Potency dependence and potential underestimation. *Risk Analysis* **12,** 115–121.

Portier, C., Hedges, J., and Hoel, D. (1986). Age-specific models of mortality and tumor onset for historical control animals in the NTP bioassay. *Cancer Research* **46,** 4372–4378.

Portier, C. and Hoel, D. (1987). Issues concerning the estimation of the TD_{50}. *Risk Analysis* **7,** 437–447.

Sawyer, C., Peto, R., Bernstein, L., and Pike, M. C. (1984). Calculation of carcinogenic potency from long-term animal carcinogenesis experiments. *Biometrics* **40,** 27–40.

Sontag, J. M., Page, N. P., and Saffiotti, U. (1975). *Guidelines for Carcinogen Bioassays in Small Rodents.* National Cancer Institute Administrative Document. Bethesda, Maryland: National Cancer Institute.

Walsh, G. R. (1975). *Methods of Optimization.* New York: Wiley.

Wartenberg, D. and Gallo, M. A. (1990). The fallacy of ranking possible carcinogen hazards using the TD_{50} (with Discussion). *Risk Analysis* **10,** 609–633.

Received April 1992; *revised August* 1992; *accepted November* 1992.