45069

Risk Analysis, Vol. 18, No. 1, 1998

Characterizing Dose–Response I: Critical Assessment of the Benchmark Dose Concept

James A. Murrell,[1] Christopher J. Portier,[2] and Richard W. Morris[1]

We present a critical assessment of the benchmark dose (BMD) method introduced by Crump[1] as an alternative method for setting a characteristic dose level for toxicant risk assessment. The no-observed-adverse-effect-level (NOAEL) method has been criticized because it does not use all of the data and because the characteristic dose level obtained depends on the dose levels and the statistical precision (sample sizes) of the study design. Defining the BMD in terms of a confidence bound on a point estimate results in a characteristic dose that also varies with the statistical precision and still depends on the study dose levels.[2] Indiscriminate choice of benchmark response level may result in a BMD that reflects little about the dose–response behavior available from using all of the data. Another concern is that the definition of the BMD for the quantal response case is different for the continuous response case. Specifically, defining the BMD for continuous data using a ratio of increased effect divided by the background response results in an arbitrary dependence on the natural background for the endpoint being studied, making comparison among endpoints less meaningful and standards more arbitrary. We define a modified benchmark dose as a point estimate using the ratio of increased effect divided by the full adverse response range which enables consistent placement of the benchmark response level and provides a BMD with a more consistent relationship to the dose–response curve shape.

KEY WORDS: Benchmark dose; dose–response; risk-assessment; NOAEL; noncancer endpoints.

1. INTRODUCTION

There has been considerable recent interest in methodologies for determining a characteristic dose level to be used in risk assessment for toxicants. Criticisms on statistical grounds of the traditional no-observed-adverse-effect-level/lowest-observed-adverse-effect level (NOAEL/LOAEL) approach to standard setting are increasingly acknowledged by biologists and there has been active debate over methods to improve upon it. While mechanistically-based mathematical modeling has received considerable attention, these techniques need more development before they can be routinely used. Instead, simpler statistical and mathematical models which relate exposure to effect have greater appeal as a means to provide an immediate, generally applicable alternative to a NOAEL/LOAEL approach. To develop an alternative to a NOAEL/LOAEL approach by such means, a method must be determined for using a model to compute an exposure level (toxicant dose) that characterizes risk. We call this exposure level the characteristic dose. In the current risk characterization practices, the characteristic dose (currently provided by the NOAEL/LOAEL) is combined with uncertainty factors and other factors to assure that the exposure standard used for regulatory purposes provides adequate safety to exposed human populations. These other factors are not addressed here. The focus of this paper is on the determination of the characteristic dose.

[1] Analytical Sciences, Inc., Durham, North Carolina.
[2] Laboratory of Computational Biology and Risk Analysis, NIEHS, Research Triangle Park, North Carolina.

0272-4332/98/0200-0013$15.00/1 © 1998 Society for Risk Analysis

Perhaps the first method for specifying a characteristic dose from a model was proposed by the U.S. Interagency Regulatory Liaison Group (IRLG).[3] This group proposed that cancer risks should be determined by fitting any appropriate dose–response model to cancer bioassay data to estimate risks over the entire range of exposure, then using a linear extrapolation of risk between a prespecified upper bound on excess risk on the model curve to the zero excess risk point. The idea was to take into account the well-documented uncertainty in the low-dose behavior of these fitted curves.[4] Excess risk for a specified low exposure dose could be calculated from this linear function. Others have proposed modifications to this method and have studied the behavior of this procedure in setting exposure standards.[5,6] One version of this method is currently proposed by the U.S. EPA for agents whose "mode-of-action" suggest linear risks at low exposure doses.[7] Recently, Gastel and Sutter[8] have attempted to specify a characteristic dose in terms of points on nonlinear dose–response curves.

The most notable of the alternatives is the benchmark dose (BMD) suggested by Crump[1,9] for noncancer risk assessment. The BMD is defined as the lower 95% confidence bound on dose which results in some prespecified level of excess risk. Recently, the benchmark dose methodology has received considerable attention. Various organizations and regulatory agencies in the U.S. and abroad have considered adoption of its use, including the U.S. EPA.[10] The benchmark dose methodology is being considered in the international arena as part of the efforts to develop a standard approach.[11]

In this paper, we examine the BMD approach from the perspective of its proposed advantages relative to the LOAEL/NOAEL approach. Specifically, we address the issues of using a model in establishing a characteristic dose, defining the characteristic dose as a lower confidence bound rather than a point estimate, and the effects of using the background response in defining excess risk or effect. To overcome shortcomings identified with the current BMD procedure, we propose a benchmark dose which is a point estimate and is based on *extra effect*, in which a change in effect over background is normalized by the range from background to maximum possible response, rather than normalized by the background response. Defined in terms of extra effect, the benchmark response can then be specified in a way such that the characteristic dose consistently characterizes the dose–response behavior described by the fitted model. Analysis given in support of the arguments uses data on noncancer endpoints of 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD) available in the literature.[12]

2. REVIEW OF THE BENCHMARK DOSE CONCEPT

2.1. Definitions and Notation

Crump defines the benchmark dose to be the 95% lower confidence limit on a dose corresponding to a fixed percent increase (relative change) in an adverse effect over the background level.[1,9] In this paper, we distinguish between a point estimate (PE) and a lower confidence bound (LCB) estimate of the benchmark dose (BMD).

The data from which benchmark dose statistics are computed derive from an experiment in which $k + 1$ groups of test animals are exposed to a specific dose level ($d_0 = 0, d_1, \ldots, d_k$) of a chemical agent (toxicant), including a zero additional dose level as a control. Adverse effects are quantified for each animal in each dose group; we will refer to the measured adverse effect on animal j in dose group i as y_{ij}; $i = 0, \ldots, k$; $j = 1, \ldots, n_i$. In order to estimate a BMD, one must specify a model relating the dose level to the average effect level for that dose group. Define $F_\theta(d)$ to be the model that describes the average adverse effect level in animals exposed to a dose d, where the model depends on a set of parameters θ. In order to use the model for estimating the BMD, the parameters in θ must be estimated from the data (using an appropriate method such as least squares or maximum likelihood estimation). The parameter estimates are denoted by $\hat{\theta}$ and the estimated model is denoted by $F_{\hat{\theta}}(d)$.

Crump[1] originally defined the BMD approach for experimental responses which are naturally expressed as proportions (e.g., proportion surviving, proportion of dams with offspring having terat, proportion of fetuses resorbed, etc.). In this case, the estimated model $F_{\hat{\theta}}(d)$ is used to compute a proportion $P(d)$ which is a number between 0 and 1 representing the risk for exposure level d. (Usually, the model $F_{\hat{\theta}}(d)$ is defined so that $P(d) = F_{\hat{\theta}}(d)$.) Mathematically, the BMD PE is defined as the dose, d_p, which satisfies the equationion

$$p = \frac{P(d_p) - P(0)}{1 - P(0)} \tag{1}$$

where $P(0)$ is the response of the control group and p (the benchmark response) is a specified level of risk. The value p, referred to as the benchmark response (BMR), is typically chosen at 0.01, 0.05, and 0.10 (1%, 5%, and 10%). The method presumes that the risk being studied can be specified as a probability ranging between $P(0)$ and 1.

The benchmark dose is then usually defined in the literature as the lower confidence bound estimated for this point estimate. The benchmark dose has been extended to experiment endpoints which are not quantal, i.e., not described in terms of proportions.[9] These endpoints are referred to in the literature as being described by continuous data. There have been several ways suggested for defining a benchmark dose in a manner similar to that in Eq. (1) for continuous data. The use of the benchmark dose concept for continuous data is discussed in more detail in Sect. 3, where we examine some of the properties of the benchmark dose as currently defined in the literature.

2.2. Findings on the Application of BMD for Studies in the Literature

That the BMD methodology is a significant contribution to risk assessment and worthy of serious consideration is supported by many authors in the literature. Kimmel[13] presented a discussion of several alternatives to the NOAEL approach, including BMD, for risk assessment of reproductive hazards. Foster *et al.*[14] discuss the adoption of the BMD method and the issues concerning its use in developmental toxicity studies and concluded that the method offers significant advantages. Several studies[15–27] have been reported which evaluate and compare the benchmark dose method to the NOAEL approach. In general, the BMD method has given similar results to NOAEL/LOAEL. However, there were instances of markedly different results in which sometimes one, sometimes the other method yielded a larger reference dose. Also, several researchers qualified their conclusions with a call for more evaluation and examination of the method.[19,21,26]

Of particular note is the review by Auton,[27] who applied the BMD method to an extensive historical database of teratology bioassays. It was found that the 5% BMD was comparable to the NOAEL for most datasets, but the 1% effective dose had large confidence intervals and hence the BMD (defined as the lower bound on the effective dose) could not be estimated accurately. It was also found that, occasionally, the BMD is higher than the highest dose tested in the study; this occurred "... when the chemical causes a small, but significant, increase in a finding that is uncommon in untreated animals."

These studies raise several important issues concerning the BMD method, e.g., instances of significant disagreement between the methods, large or incalculable confidence bounds, and applicability to a variety of endpoints (especially continuous data endpoints). In the next section, these and related issues are examined in detail.

3. CRITICAL ASSESSMENT OF BENCHMARK DOSE

3.1. Use of a Model in Computing Benchmark Dose

Computation of the benchmark dose requires the fitting of a model to the data so that a smooth inverse function is available to find the dose corresponding to a specified effect. We do not consider it a minor attribute of the benchmark dose method that dose–response data is fitted to a model in the benchmark dose approach. Nor do most of the authors that we have reviewed in the literature. This feature is always stated prominently and figures importantly in all discussions about the advocated advantages of the benchmark dose approach. Therefore, it is of great importance that this presumed advantage of using a fitted model be critically examined.

By fitting a model, the benchmark dose approach attempts to use all of the data (which contain the information about the dose–response relationship) and is not restricted to consideration of only the dose levels for which effects were measured in the study. This is a good feature, for it allows the determination of a statistic representing the character of the dose–response relationship (and thus of the potency of the toxicant). Of course, for any method using fitted models, it is important that the number and distribution of the dose levels is sufficient to yield a good estimate of the model parameters. The issue here is that the manner in which the BMD is specified can impact the degree to which "all of the data" are actually used in determining the characteristic dose.

The choice of the benchmark response (BMR) specified for the benchmark dose affects the dose region where the model is called on to make a prediction of response. It has been argued that the BMR should be "... set at the lower end of the range of response that can be detected experimentally, in order to avoid uncertainties associated with low-dose extrapolation using models that may not reflect biological realities,"[9] and yet "... large enough so that the BMD is relatively insensitive to the choice of the model."[21]

However, if the BMR is chosen so that the BMD is insensitive to the choice of the model, then curvature in the data has little effect and thus the advantage of really capturing information about the character of the dose–response behavior is not realized. To see this, consider this simple illustration using a dose–response curve

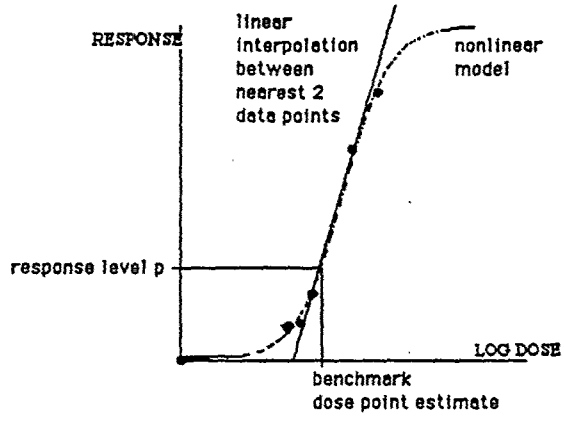

Fig. 1. A simple illustration of why choosing BMR in a linear range of response fails to take advantage of all the information available from model using all the data for the point estimate of benchmark dose.

fitted to some typical data points, shown in Fig. 1 on a log dose scale. Suppose the BMR p is chosen to lie well within the range of observed excess risks on a locally linear portion of the curve. Then there exists an experimental dose level for which the excess risk is below p and one for which the excess risk is above p. Interpolation via a straight line between these two points to find the exposure which yields the excess risk of p will yield a corresponding dose which, on average, is approximately equal to that obtained from other models fit to the same data, even nonlinear models. Hence, if p is chosen to lie so far into the range of the observed response that the curve is locally linear, then the projection of the BMR to a dose (via the model) relies almost entirely on the response at the nearest two dose levels. (This can happen frequently when p is chosen as large as 10%.) The other points are used by the fitting process to help reduce the variability in the benchmark dose statistic from noisy data, but the other points contribute virtually no information about the dose–response behavior anywhere else on the curve. That is, locating the BMD on a straight part of the curve does not provide much information about the overall dose–response relationship, and hence fails to use effectively all of the observed data in determining the BMD.

Therefore, one cannot argue that there is the advantage of using all the data, and then argue that finding a characteristic dose statistic in a linear region is best. Though it may appeal to some researchers to have a dose statistic that does not depend much on the model, we argue that the value in using a model derives from its capability to capture information about the dose–response behavior that arises from the biological mechanisms. Hence, the information carried by the model should matter, even in the case where we must resort to using merely an empirical model in the absence of an accepted mechanistic model.

A benchmark dose should be a characteristic dose statistic that characterizes the dose–response relationship by indicating a point on the curve where something interesting happens, e.g., a dramatic change in slope signifies something about the biological mechanism. The point here is that these changes can only be captured by an accurate nonlinear model, and by a choice of p such that nonlinearity affects the determination of the characteristic dose.

3.2. Use of the Lower Confidence Bound

A frequently cited advantage of the benchmark dose method is that defining it as a lower limit of a confidence interval provides industry with an incentive for performing larger studies with better sampling by rewarding it with the resulting larger benchmark dose levels.[1,9,20] While it is clear that there is a problem with the NOAEL/LOAEL approach in that the way it is used in risk assessment encourages smaller studies, there are inherent problems in defining a critical statistic of any sort in terms of an estimate of variance rather than of the mean. There are many advantages to the benchmark type of approach over the NOAEL approach, but BMD should not be defined in terms of a confidence limit. The size of the study should not bias a characteristic dose statistic in either direction. While intelligent risk assessment must be informed by the variability and uncertainty inherent in a measure, it must first have the best estimate of that measure. Once a point estimate has been found, then confidence bounds or variance estimates can be calculated and used in any of a number of various procedures in risk assessment to account for that uncertainty in standard setting (e.g., uncertainty factors, Monte Carlo simulations, fuzzy networks, etc.).

One argument that has been made is that the modeling performed as part of the benchmark dose approach, "... is not intended to provide estimates of risk levels of exposure far below the experimental dose range. Rather, the models are proposed as a means of estimating a statistical lower bound on dose associated with a predefined level of risk, that level of risk typically being in the range of 1 to 10%."[23] Attempting to deflect the charge that an empirical model is being used to make a prediction below the experimental dose region by deferring to the confidence limit is not convincing, because

Table I. 10% Extra Effect BMD Point Estimates and 95% Confidence Limits for Noncancer Endpoints of TCDD Exposure[a]

Study description	Dose regime[b]	Endpoint	10% extra effect	Lower conf. bound	Upper conf. bound
Kitchin and Woods, 1979[31]	Single, 3 days	Liver Cytochrome P-450 (total)	68.72	0.66	1337.79
female S-D rats	0.6 ng/kg	Liver Benzopyrene Hydroxalase (CYP1A1 Activity)	77.33	54.40	125.09
Abraham *et al.*, 1988[32]	Single, 7 days	Liver Cytochrome P-450 (total)	86.77	29.37	227.50
female Wistar rats	1 ng/kg	Liver EROD (CYP1A1 Activity)	157.27	108.48	244.73
Narasimhan *et al.*, 1994[33]	Single, 24 hr	Liver EROD (CYP1A1 Activity)	645.77	0.00	13925.60
female B6C3F1 mice	5 ng/kg	Liver CYP1A2 (mRNA)	391.23	0.00	9961.63
	S, 4 days	Spleen PFC/10^6 cells	29.08	0.14	91.45
Davis and Safe 1988[34]	S, 4 days	Spleen PFC/10^6 cells	359.95	39.51	1121.34[c]
C57BL/6N mice	322 ng/kg				
Mably *et al.*, 1992[35]	S, 64 ng/kg	Sperm morph.—day 120	145.6	0.00	3557.47[c]
preg. female, male offspring,		Fertility index	360.5	360.50	360.50
Holtman S-D rats		Cauda sperm count day 63	4.27	0.00	127.51[c]
		Cauda sperm count—day 120	6.83	0.00	150.50
		Cauda sperm count/g—day 120	10.10	0.00	62.31[c]
		DSP/g—day 49	0.87	0.00	4157.06[c]
		DSP/g—day 63	10.14	0.00	560.35[c]
		DSP/g—day 120	21.41	0.00	1451.92[c]
DeVito *et al.*, 1992[36]	5×/wk	liver EROD	1.08	0.06	3.98
female B6C3F1 mice	13 wk	liver Cyp1A2	0.10	0.00	4.11
VanBirgelen *et al.*, 1995[37]	daily	EROD	7.31	0.00	38.17
female S-D rats	13 wk	4OH-AA	6.86	0.00	52.66
		hepatic retinol	0.28	0.00	4.49
		hepatic retinyl-palmitate	236.90	0.00	3914.08
		plasma retinol	27.56	0.00	765.59[c]
		plasma TT4	317.17	0.00	291034[c]
		plasma FT4	51.96	0.00	1037.07[c]
VanBirgelen *et al.* 1995[38]	daily	liver T4UGT	19.55	2.49	518.57
female S-D rats	13 wk				
Maronpot *et al.*, 1993[39]	1×/2 wk	Alk. Phos. (DEN)	21.86	0.22	120.40[c]
female S-D rats	30 wk	Alk. Phos. (SAL)	33.31	0.00	74.20
		S. DeH. (DEN)	0.89	0.00	133.24

[a] Doses are in equivalent ng/kg body wt./day (multiple doses).
[b] Dose regime: period of dose administration, length of time until sacrifice.
[c] Limited by domain of search (15 times the model parameters for the MLE pt. estimate).

the confidence limit is itself an estimate which also depends on the model and is subject to all of the same assumptions and uncertainty. If the point estimate cannot be trusted outside the data range or because of a poor model fit, then its confidence bound has little meaning or relevance either.

Moreover, estimating the confidence limit (which is based on the estimated variance) is an impractical exercise for many real datasets. Table I displays results for several studies of point estimates of benchmark dose (units of micrograms/kg) from fitting a Hill equation model along with the confidence intervals (see Appendix for calculation methods). From these it is evident that the limits of the BMD often differ from the point estimate by more than 100% (sometimes by several orders of magnitude), and that often the lower limits are driven to zero. The Hill equation model is the most appropriate for data of this type, however, similar models that adequately capture the curve in the data (e.g., power law and a modified Weibull) yield similarly large gaps between point estimates and confidence bounds (results not shown).

The basic problem stems from the small sample sizes and small number of dose levels typical of most studies, combined with the nonlinearity of the model. The small samples for each dose group can inflate the estimated population variance and the small number of dose groups combined with noisy group means can lead to poor model fits (and therefore large sums of squared errors). Thus, in practice, it is not desirable to define a benchmark dose that depends on the estimated variance, such as the lower limit of a confidence interval on a

point estimate. In contrast, the point estimate itself depends more on the estimated dose–response shape, and less on the arbitrary qualities of the study.

In the case of studies with small group sample sizes and few dose groups, the NOAEL/LOAEL method typically results in exceedingly high reference doses and associated high risk. Interestingly, it is in this very case that the BMD LCB gives an unacceptably small dose (i.e., zero). In a well-designed study, the NOAEL would be very close to zero for typical dose–response curves, since a statistically significant effect would register for very low doses when there are good sampling statistics. The point here is that for both the NOAEL and the existing BMD method, the resulting risk is a nearly arbitrary reflection of the variance of the study, rather than the measured dose–response curve. When the sample variance is large, the NOAELs become larger and BMD lower confidence bounds become smaller, but both can distort the information conveyed about dose–response. In general, using the BMD PE could give a higher risk or a lower risk than that provided by a NOAEL, depending on the precision of the study, but the point is that it would be a more accurate result in terms of representing the dose–response behavior, rather than just the sampling statistics.

It is important to account for statistical uncertainty, as some have argued who advocate using the LCB as the BMD statistic. Standard statistical practice supports reporting *both* the PE and the LCB, not just the lower confidence bound. We recommend that laboratory scientists and statisticians provide both the point estimate and the confidence bounds; these are then available to the risk manager who uses a rational method for taking into account the uncertainty of the measurement. While the results from Table I certainly point to the need for better resolution studies in the future (i.e., larger samples), the database of studies already available which must be used for standard-setting would too often result in setting a standard at zero when the LCB alone is used, which is clearly unacceptable. Moreover, setting standards based on a number that reflects more about the quality of the study than it does about dose–response seems to us to be poor practice.

3.3. Standardizing the Change in Response from Background

The benchmark dose approach involves specifying the benchmark change in effect level (BMR), for example 1%, 5%, or 10% change in adverse effect or risk. In the case of quantal data models where the response represents a probability of incidence of an endpoint event (such as a particular type of tumor), it can be argued that the change in the response is a change in risk. The current definitions of BMR for continuous data do not generally represent measures of risk, but measure a numerical change in some biological quantity impacted by chemical exposure (see discussion in Kodell *et al.*).[28]

The benchmark dose is generally defined in terms of a specified change in the response from the background level. For quantal data, this change has the natural metric of a probability on a scale from 0 to 1. The difference between the probability of incidence of the endpoint at the benchmark dose and the probability for the background is already on a standardized scale. For continuous data, the importance of a change in response depends on the nature of the quantity being measured to provide some sense of the scale of the measurements. The scale is not simply a matter of units (which can be easily normalized), but of how much change in effect is really significant to public health. This differs for each quantity or endpoint measured, and can vary from study to study depending upon procedures and measurement sensitivities, species, sex, and other study design attributes. In order to obtain a statistic that conveys primarily information about the potency of the toxicant and not the particulars of the study, there must be some kind of standardization of the variation in the scale.

A variety of methods have been suggested to attempt to standardize the change from background level. The most commonly used method specifies the change in effect relative to the background response level. This *relative effect* is given by

$$E_{\text{relative}}(d) = \frac{F_\theta(d) - F_\theta(0)}{F_\theta(0)} \tag{2}$$

where $F_\theta(d)$ is the function relating the response to the dose d. The problem with relative effect is that it is highly sensitive to the level of the background response (i.e., a small effect on a small background would look more important than a similar effect on a large background). Since, the natural background response level depends arbitrarily on the particular quantity being measured, the sensitivity to change in response should have nothing to do with how much response there already is at the "background" level of dose. For example, the normal weight of an organ may be large or small depending on which organ, but a 30% change from the normal weight for one organ may be more significant to the health of the organism than a 30% change in another. Hence, E_{relative} cannot, in general, provide a standardization that permits meaningful comparisons.

Using background response as a normalization does not result in a standardized risk level for hazard ranking and risk comparisons among toxicants, among endpoints of the same toxicant, or even among studies on the same endpoint and the same toxicant. Normalizing effect difference by background can also be problematic if any responses are less than that of the control group for increasing response functions, or greater than that of the control group for decreasing functions, or if the background response level is zero.

There have been various other suggestions for standardizing the change in effect. For example, it has been suggested that the change in effect be divided by the standard deviation of the control group, or the standard deviation of an assumed distribution of the mean effect for particular dose levels.[1,21] While this method may provide a form of standardization for comparing very similar sets of data, this also does not truly standardize the change in effect, because the particular quantity measured may also vary in an arbitrary way according to the natural dispersion for that quantity. Since the standard deviation of the background can vary from endpoint and study to study in ways not related to the health importance of the effect, this method also does not standardize across a variety of endpoints and experimental conditions.

Various methods have been proposed to convert continuous data scales into the probabilistic scale by defining a cutoff response level or proportion of responses that separates a normal or no-effect region from an abnormal or adverse effect region and then quantizing the response.[1,21] Many assumptions are involved in the determination of the cut-off level and the mean response distributions. Due to the possible differences between endpoints in terms of the behavior with respect to these assumptions (as well as the uncertainty in the assumptions themselves), standardization of the type desired here cannot be achieved in this way.

Here it is proposed that a continuous quantity measurement should be scaled by the range from background response level to maximum response level (for increasing response functions). It is a biological reality that whatever the mechanism of the effect of the toxicant, there is some dose level beyond which no further change in response is seen, or is theoretically feasible. In general, there is some type of limitation or saturation phenomenon which occurs at high enough doses. Thus, in analogy to the case of quantal data for which there is a probability metric where the response reaches a maximum at one, the *extra effect* can be defined as the change in effect from background standardized by the total range of response, given by

$$E_{\text{extra}}(d) = \frac{F_\theta(d) - F_\theta(0)}{F_{\max} - F_\theta(0)} \tag{3}$$

where $F_{\max}$ is the maximum response level (the response level at saturation), i.e.,

$$F_{\max} = \lim_{d \to \infty} F_\theta(d) \tag{4}$$

Note that the total response range is *not* necessarily the response range of the *observed* responses in a study; rather, it is defined by a determination of the minimum and maximum *possible responses* according to, for example, a model equation fitted to the data. This formula is applicable to the case of quantal data, in which case $F_{\max} = 1$. The models which are frequently used for continuous, nonquantal data sets (e.g., Michaelis-Menten and Hill equation) provide an estimate of $F_{\max}$ through the estimated parameters. For most of the nonquantal datasets in the historical database available for review, whenever the study included five or more dose levels, points in the experimental range did fall near $F_{\max}$.

Dividing the change in effect by the total range of possible effects results in a quantity that is completely standardized with respect to scale. It is still arguable that the mechanism of saturation or limitation may be so different with different endpoints that comparison on the basis of this standardization may not be meaningful either. However, it is likely that a large variety of endpoints for a particular toxicant have similar limiting mechanisms, and in any case, the scale is defined in terms of the limiting possible adverse effect. The standardization thus afforded does effectively put all the endpoints on the same basis in much the same way as is done with quantal data on a probability scale (where the limiting possible health effect is occurrence of the effect with probability one). Scaling a response change by the total response range gauges the significance of the response change.

In Table II, the 10% relative effect BMD PE and its corresponding extra effect are shown for continuous data obtained from TCDD noncancer endpoint studies reported in the literature (see Appendix for description of analysis methods). Note that the percent extra effect that is actually obtained with a nominal 10% relative effect actually varies considerably with the endpoint (see also Fig. 2). The variation depends on the variation in the background response level with endpoint. If the endpoint happens to have a small background level, then the true size of the 10% relative effect is highly inflated. If the background happens to be large, the 10% relative effect is on a very insensitive scale.

Table II. 10% Relative Effect BMD Point Estimates and Corresponding Extra Effect Level: Noncancer Endpoints of TCDD Exposure (Doses Are in ng/kg Body wt. or ng/kg Body wt./day for Multiple Dose Regimes)[a]

Study description	Dose regime[b]	Endpoint	10% relative effect dose	Extra effect of 10% Rel. effect
Kitchin and Woods, 1979[31]	Single, 3 days	liver cytochrome P-450	41.31	7.82%
female S-D rats		liver benzopyrene hydroxalase	5.56	0.30%
Abraham *et al.*, 1988[32]	Single, 7 days	liver cytochrome P-450	52.46	7.19%
female Wistar rats		liver EROD	1.69	0.13%
Narasimhan *et al.* 1994[33]	Single, 1 day	liver EROD	12.37	0.24%
female B6C3F1 mice		liver Cyp1A1	2.22	0.23%
		liver Cyp1A2	257.63	2.15%
		Total Ah receptor binding	306.37	0.62%
DeVito *et al.*, 1992[36]	5×/wk	liver EROD	1.08	0.26%
female B6C3F1 mice	13 wk	liver Cyp1A2	0.10	2.47%
VanBirgelen *et al.*, 1995[37]	daily	EROD	0.27	0.19%
female S-D rats	13 wk	4OH-AA	4.51	4.25%
		plasma retinol	26.94	9.34%
VanBirgelen *et al.*, 1995[38]	daily	liver T4UGT	1.78	1.19%
female S-D rats	13 wk			

[a] Doses are in equivalent ng/kg body wt. (single dose) or ng/kg body wt./day (multiple doses).
[b] Dose regime: period of dose administration, length of time until sacrifice.

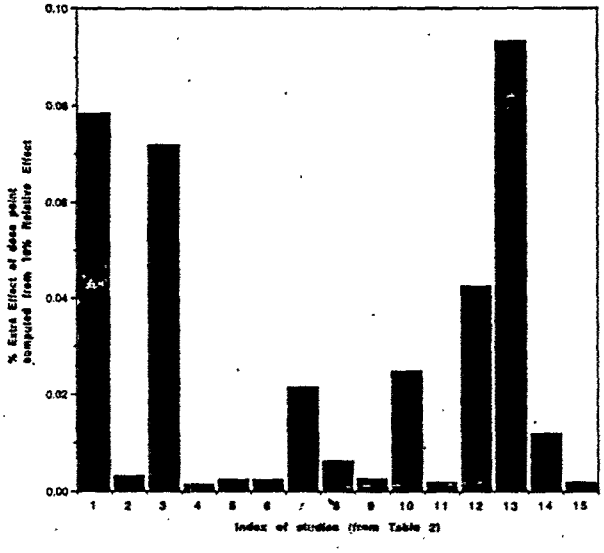

Fig. 2. Extra effect levels obtained for nominal 10% relative effect for studies in Table II.

This is illustrated in Fig. 3a and b. In Fig. 3a, the 10% relative effect and 10% extra effect are fairly close on the dose–response curve. However, in Fig. 3b, the 10% relative effect benchmark dose occurs at a much smaller dose and lower on the curve. In fact, as occurs with many cases, the benchmark dose occurs in the very, very small dose range; there are no dose group data points between the BMD and the control group point (i.e., it is a "low-dose extrapolation"). Recalling the earlier discussion about selection of the BMR, the evidence suggests that the 10% relative effect may not place the BMD in the region that has been characterized as desirable by some proponents of benchmark dose.

Some may try to argue that the size of the background level is not really arbitrary and can reasonably be expected to track with the sensitivity of the scale of the measured quantity. Figure 4 shows that the log of the response range increases linearly with the log of the background response level. Note however, that the relationship is not linear in the original scale of the background and total response range, which would be required in order for standardization by a simple division to be effective. Also, there is considerable noise even in the log–log relationship. Hence, dividing by background rather than range gives different results than dividing by response range in general, and they do not track each other very well.

The usual discussions in the literature about models address the fact that it is the low end of the dose range that is of interest in toxicology for the purposes of regulation. We propose nothing here that runs counter to that inescapable reality. Our point in this section is that increments in response due to increments of dose (at any part of the dose range) should be judged in severity according to the scale of full range of possible adverse response, rather than according to the relative scale of the background level of response. This is because, as we have shown, the background response level can have a

quite arbitrary relationship to the severity of the adverse response, whereas the size of the response range from background to maximum possible response should reasonably be expected to provide a meaningful scale for the observed effects.

4. A BENCHMARK DOSE POINT ESTIMATE BASED ON EXTRA RISK

Considering the issues discussed in Sect. 3, a suggested improvement over the currently defined benchmark dose method is to use the effective dose that results in a specified percent change in the *extra effect*. That is, the point estimate of the $p \times 100\%$ benchmark dose is the dose d_p which satisfies:

$$p = \frac{F_\theta(d) - b}{v} \tag{5}$$

where $b = F_\theta(0)$ is the background response and $v = F_{max} - F_\theta(0)$ is the range of response (e.g., from background to maximum response for an increasing dose–response relationship). Note that this same definition applies to the case of quantal data as well as the case for continuous data.

The point estimates of effective dose for 5% extra effect are compared to those of 5% relative effect in Table III. Sometimes they agree well, but often the quantities are radically different. Where they are similar is where the background response $b = F_\theta(0)$ happens to be close to the value of the total response range $v = F_{max} - F_\theta(0)$; that this should be the case is evident from examining Eqs. (2), (3), and (5).

We argue that the shape of the dose–response curve is crucial to the determination of the importance of an effect and the consequent characterization of risk.[30] As seen in Table III, the extra effect doses always have a fixed relationship to the dose–response characteristics as indicated by the fitted model parameters [using the Hill equation $F_\theta(d) = b + vd^n/(k^n + d^n)$; see Appendix]. By the Hill equation, the reciprocal of the log of the ratio of the dose to k should be linearly related to the exponent parameter n. Figure 5 shows this relationship using the benchmark dose point estimate (5% effective dose). Since k and n characterize the shape of the dose–response curve for the Hill equation, it is hoped that the benchmark dose statistic used would consistently reflect the characteristic of the dose–response relationship. The extra effect benchmark dose statistic (normalization of effect by the response range) shows a consistent relationship among BMD, the k parameter and the n parameter. Due to the fact that a fixed percentage in terms of relative effect actually varies considerably in terms of extra effect (as shown above), the effective dose (benchmark dose) obtained from relative effect does not have a fixed relationship to the shape of the dose curve. Figure 5 shows an erratic and unpredictable relationship between the characteristic dose–response curve and the relative effect benchmark dose points. It is clear that the

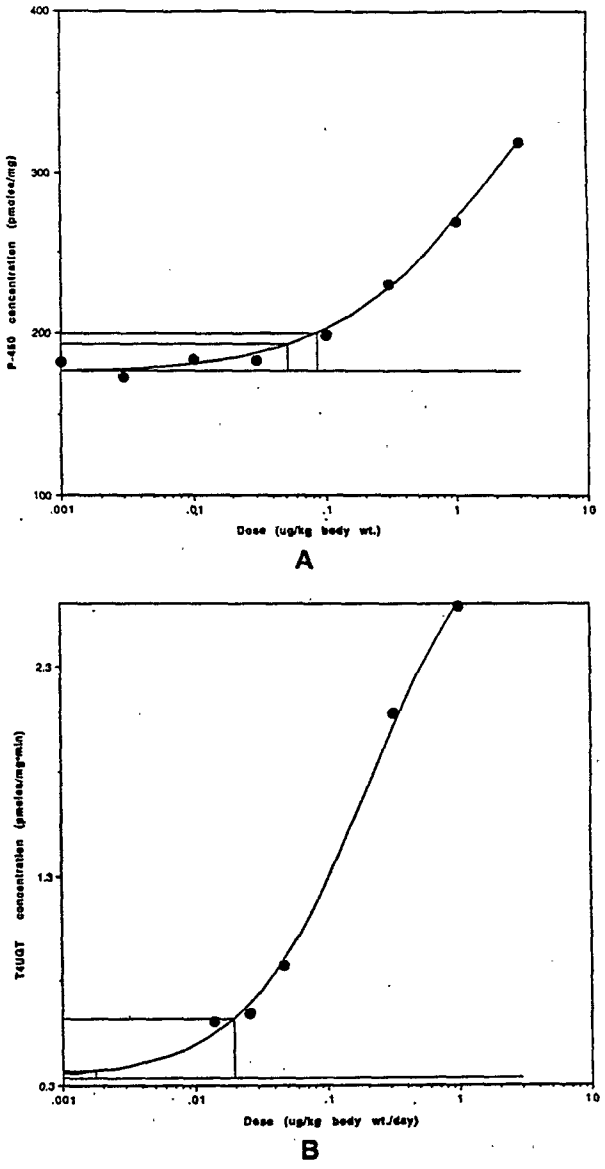

Fig. 3. 10% Relative effect point (lower pt.) and 10% Extra effect point (upper) on Hill equation fitted curve. (a) Liver P450 response to dioxin (Abraham *et al.*, 1988).[32] (b) Liver T4UGT response to dioxin (Van Birgelen *et al.*, 1995a).[37]

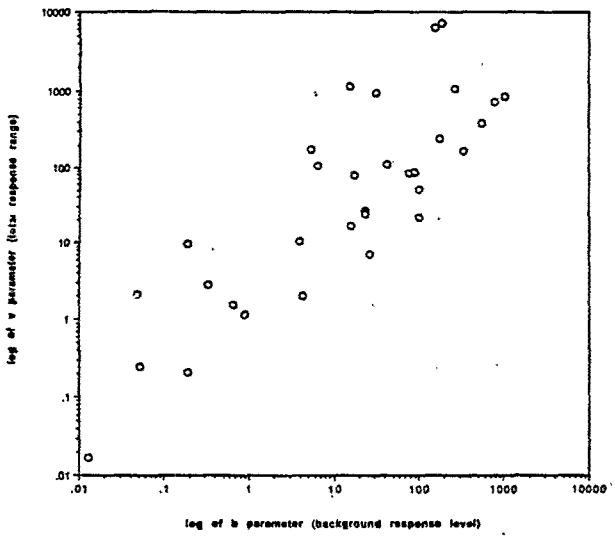

Fig. 4. Response range plotted against background level, log–log scale (data from the all the studies in Table II except Mably, 1992).[35]

definition of BMD for continuous data in terms of response relative to background is not consistently related to dose–response, and therefore cannot give a consistent characterization of health risk.

Since the benchmark dose obtained using extra effect has a fixed relationship to the estimated dose–response pattern, the specified benchmark response level can be chosen so that the BMD always falls at a place on the dose–response curve where a characteristic nonlinearity of interest occurs, such as a region on the dose–response curve where the response is rapidly increasing with increased dose.

The advantage of extra effect BMD over the relative effect BMD is that all continuous data endpoints are normalized to the same scale; hence, it is truly an indicator of the potency of the toxicant independent of the natural background level for the measured response and independent of the sensitivity of the response to dose level.

5. SUMMARY AND CONCLUSIONS

As the benchmark dose methodology gains greater acceptance and gains momentum toward being adopted as a risk assessment alternative to the NOAEL/LOAEL approach, it is important to provide a critical assessment of its strengths and weaknesses, which is the objective of this report. It is certainly an advantage to use a model-based method in order to use more of the information in the data. However, the key information in dose–response modeling is in the nonlinear characteristic behavior of a toxicant for particular endpoints. The procedure used to extract a dose level to be employed in standard setting should reflect this curvature.

Several aspects of the current benchmark dose procedure described in the literature interfere with its successfully characterizing the dose–response relationship. First, the benchmark response level might be chosen on a locally linear region, which does result in the ostensibly desirable "model insensitivity," but as we have shown, it also results in a characteristic dose statistic that conveys little information about nonlinear dose–response behavior, thus obviating the supposed advantage of fitting a model to the data. We argue that the benchmark response level should be chosen to pick up information about dose–response behavior at a characteristic nonlinear region of the curve to truly take advantage of the information provided by fitting a model to the data.

Second, the use of the lower confidence bound estimate rather than a point estimate leads to a characteristic dose that reflects more the lack of information in typical studies than it does the trend information that is available in the data. We have shown that for the continuous-data endpoints of TCDD analyzed in this study, the lower confidence bound can vary widely according to the precision of the study and that the use of the lower confidence would lead to a dose standard at or near zero in many cases. The results show that a standard based solely on the lower confidence bound would have a greater dependence on the precision of available studies than on measured dose–response, again obviating any advantage to performing dose–response modeling.

Third, we have shown that for continuous data endpoints, most methods for standardizing the response scale for defining excess effect can result in BMDs which are not comparable. In particular, the results show that the widely-used method of scaling by background level yields widely differing scale sensitivities. Using the background response level to scale change in effect produces a BMD that depends on whatever happens to be the natural background response level (which can be different among endpoints or experimental conditions even when the mechanism of toxicity is similar). We have demonstrated with analysis of data that this relative effect BMD does not track with the extra effect BMD with scaling based on the range of response from background to maximum. We argue that extra effect scales the severity of adverse effect better and show that it tracks better with model parameters indicating dose–response behavior.

Table III. Comparison of Relative Effect BMD to Extra Effect BMD for Noncancer Endpoints of TCDD Exposure (Doses are in ng/kg Body wt. or ng/kg Body wt./day for multiple dose regimes)[a]

Study description	Dose regime[b]	Endpoint	5% relative effect BMD	5% extra effect BMD
Kitchin and Woods, 1979[31]	Single,	liver cytochrome P-450	10.31	16.75
[illegible] S-D rats	3 days	liver benzopyrene hydroxalase	3.34	44.76
[illegible] *et al.*, 1988[32]	Single,	[illegible] cytochrome P-450	18.94	30.64
[illegible] Wistar rats	7 days	liv[illegible]ROD	0.83	73.04
Narasimhan *et al.*, 1994[33]	Single,	liver EROD	6.04	298.77
female B6C3F1 mice	1 day	liver Cyp1A1	1.11	51.36
		liver Cyp1A2	214.83	322.62
		Total Ah receptor binding	255.66	534.21
	S, 4 days	Spleen PFC/10^6 cells	25.56	22.68
Davis and Safe 1988[34]	S, 4 days	Spleen PFC/10^6 cells	320.34	305.75
C57BL/6N mice				
Mably *et al.*, 1992[35]	Single	Sperm morph.—day 120	148.76	121.89
preg: female, male offspring,		Fertility index	379.33	350.53
Holtman S-D rats		Cauda sperm count—day 63	2.94	2.02
		Cauda sperm count—day 120	6.79	3.49
		Cauda sperm count/g—day 120	15.27	6.56
		DSP/g—day 49	1.33	13.84
		DSP/g—day 63	21.31	0.16
		DSP/g—day 120	28.19	6.95
DeVito *et al.*, 1992[36]	5×/wk	liver EROD	0.65	10.00
female B6C3F1 mice	13 wk	liver Cyp1A2	0.04	0.27
VanBirgelen *et al.* 1995[37]	daily	EROD	0.16	3.99
female S-D rats	13 wk	4OH-AA	3.25	4.87
		hepatic retinol	0.09	0.09
		hepatic retinyl-palmitate	227.16	224.63
		plasma retinol	21.54	22.02
		plasma TT4	19.06	96.63
		plasma FT4	27.46	24.88
VanBirgelen *et al.* 1995[38]	daily	liver T4UGT	0.84	8.73
female S-D rats	13 wk			
Maronpot *et al.*, 1993[39]	1×/2 wk	Alk. Phos. (DEN)	16.61	17.26
female S-D rats	30 wk	Alk. Phos. (SAL)	30.54	28.20
		ALT	0.07	0.09
		Triglycerides	14.36	8.70
		S. DeH. (DEN)	0.26	4.12
		S. DeH. (SAL)	31.15	28.12
		Liver labeling index	28.59	30.77
Tritscher *et al.*, 1993[40]	1×/2 wk	Max. EGF receptor	6.14	4.50
female S-D rats	30 wk	Avg. Conc 1A2	0.014	2.90
		Mean focal volume	29.11	29.52

[a] Doses are in equivalent ng/kg body wt. (single dose) or ng/kg body wt./day (multiple doses).
[b] Dose regime: period of dose administration, length of time until sacrifice.

We put forward that the benchmark dose methodology could be a sound one, for both quantal and continuous data, if it is defined as a point estimate (accompanied by confidence limits to indicate the quality of the estimate) and put in terms of *extra effect*. This proposed modification of the benchmark dose method has several key advantages. Not using a lower confidence bound to determine the characteristic dose avoids the arbitrary dependence of a dose standard on the precision of studies and the possibly absurd result of zero dose. Defining in terms of extra effect for continuous data puts the scale on a basis similar to that of the proportional scale for quantal incidence data cases. Using this method, the benchmark response level can be consistently defined and applied so that it falls in the region in the dose–response curve where important (potentially

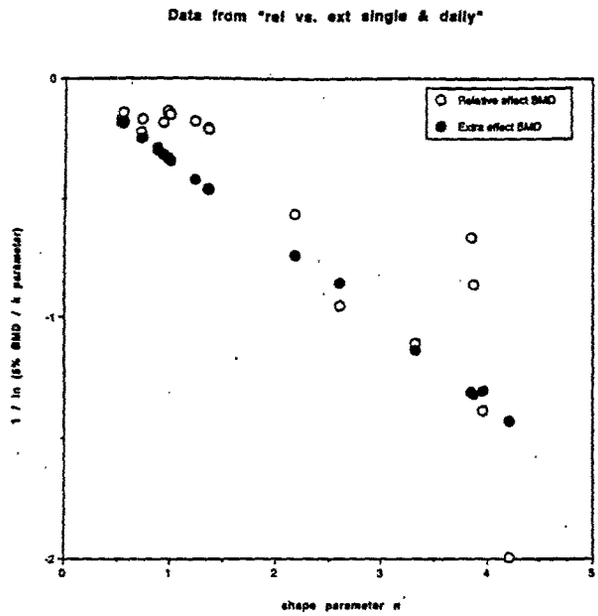

Fig. 5. Relationship between BMD/k and n. (data from studies in Table II having a single dose or daily dose regimen).

biologically important) dose–response behavior occurs that characterizes the behavior of the toxicant. We have also shown that the BMD defined in terms of extra effect has a consistent relationship with model parameters characterizing dose–response shape.

APPENDIX: CALCULATION METHODS

A1. Fitting the Models to the Data

Nonlinear regression of response on dose was performed on noncancer endpoints using the NLIN procedure of SAS/STAT 6.10. Least squares estimates were obtained using the Marquardt method.[41] In each data set, a continuous response is expressed in the original units. An intercept term was included in all models to estimate the response level at zero administered dose (background level).

Depending on the availability of the data, the regression analyses used the raw data (with a data point for each animal in the study), or the mean of each dose group weighted by the inverse of the standard error, or simply the unweighted means. In some cases, analyses were not reported if there were no trend discernible by examining a plot of the data. Some datasets were excluded because noisy data in combination with a trend having a very small slope resulted in fitting noise rather than trend. In practice, five or more dose levels (including control group) were needed to yield reasonable fits.

For continuous dose–response data of this type, the Michaelis-Menten model or the Hill equation model are typically used. For endpoints which are biochemical markers of dioxin activity, there is some mechanistic justification of these models in terms of dioxin binding to a limited supply of cell Ah receptors. The Hill equation model is chosen here to model the dose–response relationship because it is more general and can flexibly capture the steepness of the dose–response shape with its exponent parameter. The Hill equation model with an intercept is given as:

$$E(y) = b + \frac{v d^n}{k^n + d^n} \qquad \text{(for increasing D-R curve)}$$

$$E(y) = b - \frac{v d^n}{k^n + d^n} \qquad \text{(for decreasing D-R curve)}$$

where y is the response, $E(y)$ is the expected value of the response, d is the dose level, b is the intercept term (estimates the background response), n is the Hill exponent and is called the shape parameter because it characterizes the curvature of the dose–response, v is the maximum response above background ("V_{max}" of the Hill equation), and k is a characteristic quantity that can be interpreted as the dose at which the response is half the maximum response above background (the K_d, Hill coefficient).

For the Hill model, the relative effect benchmark dose is given by:

$$d^* = k\left[\frac{bp}{v - bp}\right]^{1/n}$$

For the Hill equation model, the extra effect benchmark dose is given by:

$$d^* = k\left[\frac{p}{1 - p}\right]^{1/n}$$

Note that for extra effect definition of benchmark dose, BMD depends only on the dose–response curve shape parameter n (which is a dimensionless, scaleless parameter) and k, which is units of dose and acts as a location parameter that fixes the location of the curve.

A2. Computing Confidence Intervals for Benchmark Dose

Constrained optimization was used to find maximum and minimum dose d satisfying equations

$$p = \frac{F_\theta(d) - F_\theta(0)}{F_{max} - F_\theta(0)} \quad \text{(A1)}$$

$$\frac{S(\theta) - S(\theta^*)}{S(\theta^*)} \frac{dfe}{dfm} \leq F(95\%, dfm, dfe) \quad \text{(A2)}$$

for all [illegible] quation parameters on the confidence surface determined by sums of squares satisfying a 95% F statistic (with degrees of freedom according to the number of data points in the study and number of model parameters), where

$S(\theta) = \sum_{i=1}^{N} (Y_i(d_i) - y(\theta,d_i))^2$
θ = vector of model parameters
θ^* = maximum likelihood estimate (MLE) of the model parameters
N = number of observations
dfm = degrees of freedom for the model (number of model parameters)
dfe = degrees of freedom for the errors ($N - dfm$)
$Y_i(d_i)$ = observed response at ith experimental dose d_i
$y(d,\theta)$ = response predicted by the model at dose d using parameters θ

The optimization algorithm used is FORTRAN IMSL routine NCONF, which is based on Schittkowski's NLPQL algorithm;[42] it uses a successive quadratic programming method to solve a nonlinear programming problem. The parameter space is restricted to nonnegative parameter values (nonzero for k and n); to avoid the numerical difficulties of limited precision, an upper limited of 15 times greater than the parameter point estimates is also imposed. Initial values for the optimization routine are obtained using a grid search.

REFERENCES

1. K. S. Crump, "A Method For Determining Allowable Daily Intakes," *Fundament. Appl. Toxicol.* 4, 854–871 (1984).
2. E. A. Weller, P. J. Catalono, and P. L. Williams, "Implications of Developmental Toxicity Study Design for Quantitative Risk Assessment," *Risk Anal.* 15(5) (1995).
3. International Regulatory Liason Group (IRLG), Scientific Bases for Identification of Potential Carcinogens and Estimation of Risks, Federal Register (July 6, 1979).
4. D. Hoel, "Mathematical Dose Response Models and Their Application to Risk Estimation," Proceedings of the Workshop in Quantitative Estimation of Risk to Human Health from Chemicals, Rome, Italy (July 12–16, 1982).
5. D. Krewski, D. Gaylor, and M. Szyskowicz, "A Model-Free Approach to Low-Dose Extrapolation," *Environ. Health Perspect.* 90, 279–285 (1991).
6. D. W. Gaylor and R. L. Kodell, "Linear Interpolation Algorithm for Low Dose Risk Assessment of Toxic Substances," *J. Environ. Pathol. Toxicol.* 4, 305–312 (1983).
7. U.S. Environmental Protection Agency (EPA), Guidelines for Carcinogenic Risk, Federal Register 5 (185), 33992-34003 (1996).
8. J. A. Gastel and T. R. Sutter, "Biologically Bounded Risk Assessment for Receptor-Mediated Nongenotxic Carcinogens," *Regul. Toxicol. Pharmacol.* 22, 273–292 (1995).
9. K. S. Crump, "Calculation of Benchmark Doses from Continuous Data," *Risk Anal.* 15(1), 79–89 (1995).
10. U.S. Environmental Protection Agency (EPA), Benchmark Dose Technical Guidance Document, EPA/600/P-96/002 (1995).
11. International Program on Ch[illegible]afety (IPCS) [illegible] Workshop on the Harmo[illegible] of Risk [illegible] for [illegible]ity and Mutagenicity (Germ Cells), International Program on Chemical Safety, Report PCS-C18/445/22, World Health Organization, Geneva, Switzerland (1995).
12. U.S. Environmental Protection Agency (EPA), Health Reassessment Document for 2,3,7,8-Tetrachlorodibenzo-*p*-Dioxin (TCDD) and Related Compounds (in press).
13. C. A. Kimmel, "Approaches to Evaluating Reproductive Hazards and Risks," *Environ. Health Perspect.* 101(2), 137–143 (1993).
14. P. M. D. Foster and T. R. Auton, "Application of Benchmark Dose Risk Assessment Methodology to Developmental Toxicity—An Industrial View," *Toxicol. Lett.* 82-3, 555–559 (1995).
15. J. L. Larson, M. V. Templin, D. C. Wolf, K. C. Jamison, J. R. Leininger, S. A. Mery, K. T. Morgan, B. A. Wong, R. B. Conolly, and B. E. Butterworth, "90-Day Chloroform Inhalation Study in Female and Male B6C3F(1) Mice–Implications for Cancer Risk Assessment," *Fundament. Appl. Toxicol.* 30(1), 118–137 (1996).
16. F. J. Murray, "A Human Health Risk Assessment of Boron (Boric-Acid and Borax) in Drinking-Water," *Regul. Toxicol. Pharmacol.* 22(3), 221–230 (1995).
17. C. A. Kimmel, R. J. Kavlock, B. C. Allen, and E. M. Faustman, "The Application of Benchmark Dose Methodology to Data from Prenatal Development Toxicity Studies," *Toxicol. Lett.* 82-3, 549–554 (1995).
18. P. A. Maisch, D. M. Proctor, and B. L. Finley, "Estimation of a Chromium Inhalation Reference Concentration Using the Benchmark Dose Method—A Case-Study," *Regul. Toxicol. Pharmacol.* 20(1), 58–82 (1994).
19. W. Pease, J. Vandenberg, and K. Hooper, "Comparing Alternative Approaches to Establishing Regulatory Levels for Reproductive Toxicants—DBCP as a Case-Study," *Environ. Health Perspect.* 91, 141–155 (1991).
20. M. Haag-Gronlund, R. Franssonsteen, and K. Victorin, "Application of the Benchmark Method to Risk Assessment of Trichloroethene," *Regul. Toxicol. Pharmacol.* 21(2), 261–269 (1995).
21. W. Slikker, K. S. Crump, M. E. Andersen, and D. Bellinger, "Biologically-Based, Quantitative Risk Assessment of Neurotoxicants," *Fundament. Appl. Toxicol.* 29(1), 18–30 (1996).
22. R. J. Kavlock, B. C. Allen, E. M. Faustman, and C. A. Kimmel, "Dose–Response Assessments for Developmental Toxicity 4. Benchmark Doses for Fetal Weight Changes," *Fundament. Appl. Toxicol.* 26(2), 211–222 (1995).
23. B. C. Allen, R. J. Kavlock, C. A. Kimmel, and E. M. Faustman, "Dose–Response Assessment for Developmental Toxicity 2: Comparison of Generic Benchmark Dose Estimates with No Observed Adverse Effect Levels," *Fundament. Appl. Toxicol.* 23(4), 487–495 (1994).
24. B. C. Allen, R. J. Kavlock, C. A. Kimmel, and E. M. Faustman, "Dose–Response Assessment for Developmental Toxicity 3: Statistical-Models," *Fundament. Appl. Toxicol.* 23(4), 496–509 (1994).
25. E. M. Faustman, B. C. Allen, R. J. Kavlock, and C. A. Kimmel, "Dose–Response Assessment for Developmental Toxicity. 1. Characterization of Database and Determination of No Observed Adverse Effect Levels," *Fundament. Appl. Toxicol.* 23(4), 478–486 (1994).
26. D. G. Barnes, G. P. Daston, J. S. Evans, A. M. Jarabek, R. J. Kavlock, C. A. Kimmel, C. Park, and H. L. Spitzer, "Benchmark Dose Workshop—Criteria for Use of a Benchmark Dose to Esti-

mate a Reference Dose," *Regul. Toxicol. Pharmacol.* **21**(2), 296–306 (1995).

27. T. R. Auton, "Calculation of Benchmark Doses from Teratology Data," *Regul. Toxicol. Pharmacol.* **19**(2), 152–167 (1994).
28. R. L. Kodell, J. J. Chen, and D. W. Gaylor, "Neurotoxicity Modeling for Risk Assessment," *Regul. Toxicol. Pharmacol.* **22**(1), 24–29 (1995).
29. D. W. Gaylor and W. Slikker, Jr., "Risk Assessment for Neurotoxic Effects," *NeuroToxicology* **11**, 211–218 (1990).
30. K. L. Meier, A. J. Bailer, and C. J. Portier, "A Measure of Tumorigenic Potency Incorporating Dose–Response Shape," *Biometrics* **49**, 917–926 (1993).
31. K. T. Kitchin and J. S. Woods, "2,3,7,8-Tetrachlorodibenzo-*p*-dioxin (TCDD) Effects on Hepatic Microsomal Cytochrome P-448-Mediated Enzyme Activities," *Toxicol. Appl. Pharmacol.* **47**, 537–546 (1979).
32. K. Abraham, R. Krowke, and D. Neubert, "Pharmacokinetics and Biological Activity of 2,3,7,8-tetrachlordibenzo-*p*-dioxin: 1. Dose-Dependent Tissue Distribution and Induction of Hepatic Ethoxyresorufin O-Deethylase in Rats Following a Single Injection," *Arch. Toxicol.* **62**, 359–368 (1988).
33. T. R. Narasimhan, A. Craig, L. Arellano, N. Harper, L. Howie, M. Menache, L. Birnbaum, and S. Safe, "Relative Sensitivities of 2,3,7,8-Tetrachlorodibenzo-*p*-dioxin-Induced Cyp1a-1 and Cyp1a-2 Gene Expression and Immunotoxicity in Female B6C3F1 Mice," *Fundament. Appl. Toxicol.* **23**, 598–607 (1994).
34. D. Davis and S. Safe, "Immunosuppressive Activities of Polychlorinated Dibenzofuran Congeners: Quantitative Structure-Activity Relationships and Interactive Effects," *Toxicol. Appl. Pharmacol.* **94**, 141–149 (1988).
35. T. A. Mably, D. L. Bjerke, R. W. Moore, A. Gendron-Fitxpatrick, and R. E. Peterson, "*In Utero* and Lactational Exposure of Male Rats to 2,3,7,8-Tetrachlorodibenzo-*p*-dioxin: 3. Effects on Spermatogenesis and Reproductive Capability," *Toxicol. Appl. Pharmacol.* **114**, 118–126 (1992).
36. M. J. DeVito, T. Thomas, E. Martin, T. Umbreit, and M. A. Gallo, "Antiestrogenic Action of 2,3,7,8-Tetrachlorodibenzo-*p*-dioxin: Tissue Specific Regulation of Estrogen Receptor in CD1 Mice," *Toxicol. Appl. Pharmacol.* **113**, 284–292 (1992).
37. A. P. J. M. Van Birgelen, J. Van Der Kolk, K. M. Fase, I. Bol, H. Poiger, A. Brouwer, and M. Van Den Berg, "Subchronic Dose–Response Study of 2,3,7,8-Tetrachlorodibenzo-*p*-dioxin in Female Sprague-Dawley Rats," *Toxicol. Appl. Pharmacol.* **132**, 1–13 (1995).
38. A. P. J. M. Van Birgelen, E. A. Smit, I. M. Kampen, C. N. Goeneveld, K. M. Fase. J. Van Der Kolk, H. Poiger, M. Van Den Berg, J. H. Koeman, and A. Brouwer, "Subchronic Effects of 2,3,7,8-Tetrachlorodibenzo-*p*-dioxin or PCBs on Thyroid Hormone Metabolism: Use in Risk Assessment," *Eur. J. Pharmacol.; Environ. Toxicol. Pharmacol. Sect.* **293**, 77–85 (1995).
39. R. R. Maronpot, J. F. Foley, K. Takahashi, T. Goldsworthy, G. Clark, A. Tritscher, C. Portier, and G. Lucier, "Dose Response for TCDD Promotion of Hepatocarcinogenesis in Rats Initiated with DEN: Histologic, Biochemical, and Cell Proliferation End Points," *Environ. Health Perspect.* **101**(7), 634–643 (1993).
40. A. M. Tritscher, J. A. Goldstein, C. J. Portier, Z. McCoy, G. Clark, and G. Lucier, "Dose–Response Relationships for Chronic Exposure to 2,3,7,8-Tetrachlorodibenzo-*p*-dioxin in a Rat Tumor Promotion Model: Quantification and Immunolocalization of CYPIA1 and CYPIA2 in the Liver," *Cancer Res.* **52**, 3426–3442 (1993).
41. D. W. Marquardt, "An Algorithm for Least Squares Estimation of Nonlinear Parameters," *J. Soc. Indust. Appl. Math.* **11**, 431–441 (1963).
42. K. Schittkowski, "NLPQL: A FORTRAN Subroutine Solving Constrained Nonlinear Programming Problems," *Ann. Operat. Res.* **5**, 485–500 (1986).