Review of Noncancer Risk Assessment:

Application of Benchmark Dose Methods

Prepared for the

Commission on Risk Assessment and Risk Management

by

Elaine M. Faustman, Ph.D., DABT
Professor and Associate Chairperson
Department of Environmental Health
School of Public Health and Community Medicine
University of Washington
Seattle, WA

June 1996

Review of Noncancer Risk Assessment: Applications of Benchmark Dose Methods

Elaine M. Faustman, PhD, DABT

Professor and Associate Chairperson

Department of Environmental Health

School of Public Health and Community Medicine

University of Washington

4225 Roosevelt Way NE, Suite 100

Seattle, WA 98105-6099

Phone: (206) 685-2269

Fax: (206) 685-4696

email: efaustman@dehpost.sphcm.washington.edu

Abstract

The overall goal of this project is to evaluate risk assessment methods traditionally used for noncancer health risks and to compare them with new approaches. Initially this report gives a brief economic rationale for the impact of prevention of noncancer health effects. This is shown using figures for years of potential life lost which reveal that noncancer health effects, such as birth defects, are on the same national economic magnitude as cancer and heart disease. Traditional approaches for assessing these noncancer risks are discussed. These methods include identification of no observed adverse effect levels (NOAELs). Reference dose (RfD) calculation or setting of acceptable daily intake (ADI) values is achieved by dividing the NOAEL values by uncertainty and/or modifying factors. These factors represent a default approach to

account for animal-to-human and average-to-sensitive population extrapolation or extrapolation from inadequately designed experiments. In the case where all doses tested produced a response, then the use of a lowest observed adverse effect level (LOAEL) is described and application of an additional 10-fold factor is discussed. These traditional approaches are compared to benchmark dose methods where a specific effect level is identified using a curve-fitting procedure in the range of biologically observable data. Confidence limits are generated around this dose and the benchmark dose is set at the lower confidence limit producing an x% change in response (BMDx). The BMDx can be used to calculate a reference dose using a similar default safety factor approach as for the NOAEL calculations.

Applications of the BMD method are given for noncancer toxicity endpoints. Although the majority of applications of this approach are for developmental toxicity endpoints, it has also been applied for reproductive toxicity, neurotoxicity, and cancer endpoints. Discussion of these applications both within the US and in the international community is given. The most thorough evaluation of this method is for developmental toxicity and is published in a series of 4 papers and technical documents by Faustman, Allen, Kavlock, and Kimmel (Faustman *et al.*, 1994; Allen, 1994a, b; Kavlock *et al.*, 1995) analyzing over 1,825 experimental endpoints. These evaluations show that the benchmark method offers an alternative to traditional NOAEL approaches that are in general no more conservative than the NOAEL approach and which include a confidence limit calculation. The authors identify a log-logistic model approach for BMD estimation for developmental toxicity as having several advantages, and they show that BMD_5 values generated using this model are similar to both continuous and quantal NOAEL values (without confidence limits). These authors also show that traditional safety factor approaches used for RfD calculation based on LOAEL values are overly conservative and that, rather than a 10-fold factor, a factor of 5 is more appropriate. These studies show that the NOAEL values are not "risk free" but represent effect levels

ranging from below 5% up to 20% effect. These observations are consistent with previous observations where Leisenring and Ryan (1992) determined that the risk associated with NOAEL values from dose response curves can range between 3 and 21%. Using developmental toxicity data, Gaylor (1992) shows that for 25% of the cases evaluated, the NOAEL responses were between 1 and 4.5% response (MLE response). This illustrates an important advantage of BMD approaches in that a regulatory limit can be consistently set at a given response level rather than being dictated by study design. The benchmark dose method rewards adequately designed experiments by setting higher BMD values, which is in direct contrast to the NOAEL approach. Using curve-fitting procedures, the calculation of RfD values is no longer constrained to be one of the experimental doses tested. BMD methodology will allow for easy transition to truly biologically based dose response modeling when such models are developed. This review discusses several new areas of research on this topic. In summary, both research and philosophical advantages of BMD approaches are given in this report.

Introduction

The overall goal of this project is to evaluate risk assessment methods traditionally used for noncancer health risks and to compare them with new approaches. To put the social impact of noncancer health effects into perspective, this paper refers to economic costs. We use developmental toxicity as an example for showing the financial costs of a noncancer effect.

Birth defects are the leading cause of infant mortality and the fifth leading cause of years of potential life lost (YPLL) in the United States. Substantial resources have been allocated to care for persons with birth defects, but the economic impact of these resources had not been calculated until a recent article in Morbidity and Mortality Weekly Report (MMWR, 1995). That article used a human-capital approach. The

economic costs of medical, developmental, and special-education services were calculated and added to the indirect costs of loss of work and household productivity attributable to premature mortality and morbidity in persons with any of 18 adverse developmental outcomes. The estimates were based on cerebral palsy and 17 of the clinically most important structural birth defects observed in the United States. Table 1 shows the list of defects included in the analysis and the estimated economic costs associated with each.

The economic cost associated with the selected conditions for 1992 was $8 billion. Such figures do not include consideration of noneconomic factors, such as impact on families and the psychosocial costs of illness, nor do they include developmental defects besides the 18 listed outcomes. Thus, the costs reported were low-end estimates.

Years of potential life lost can also be evaluated for adverse developmental endpoints. Over 1.6 million YPLL (based on an average life of 65 years) were estimated as due to developmental defects, such as congenital anomalies, prematurity, and sudden infant death syndrome. That figure is close to that calculated for malignancies and neoplasms (1.8 million YPLL) and surpasses that for heart conditions (1.5 million YPLL) (MMWR, 1990; NAS, 1996). Those figures should help to place developmental disabilities and other noncancer health endpoints in the same arena of public health concern as cancer.

Estimates suggest that environmental factors play a role in causing 10-17% of all birth defects. Over 65% have unknown causes (Faustman *et al.*, 1995). The economic costs and YPLL figures suggest the large potential economic benefit of public health measures that would identify and prevent non cancer effects.

Traditional Methods for Assessing Risks of Noncancer Endpoints

The standard procedure for assessing noncancer risks associated with hazardous compounds has been to use a no observed adverse effect level (NOAEL) approach in which a no-effect level is identified and an uncertainty factor is applied to it to estimate a dose for humans that is below a presumed threshold and which represents an acceptably safe exposure level. The uncertainty factor is intended to account for variability in response to a given level of exposure both within and across species. Some of the serious drawbacks of such an approach have been highlighted (Crump, 1984; Kimmel and Gaylor, 1988) and one of the purposes of this report is to delineate and illustrate these limitations.

Traditionally, NOAELs are used to establish human permissible exposure levels for noncarcinogens. Examples of these allowable exposure levels are acceptable daily intakes (ADIs), threshold limit values (TLVs), and reference doses and concentrations (RfDs and RfCs). ADIs and RfDs are derived from a NOAEL by applying safety factors: uncertainty factors (UFs) and modifying factors (MFs):

$$RfD = NOAEL / [(UF)\ (MF)]$$

$$ADI = NOAEL / [(UF)\ (MF)]$$

The safety factors allow for intraspecies and interspecies (animal to human) variation. Default UFs of 10 are assigned when relevant research-based information is missing. MFs are used to adjust the UFs if data on pharmacokinetics, pharmacodynamics, or mechanisms is available to evaluate the relevance of animal information for human responses. If a NOAEL has not been determined from the available studies, an additional factor of 10 is applied to the lowest observed adverse effect level (LOAEL) to get a value that is more comparable to a NOAEL.

Criticisms of the NOAEL approach include the following issues:

(1) The NOAEL by definition is one of the experimental doses tested. Because of that constraint, the rest of the dose response relationship is largely ignored.

(2) The NOAEL approach does not identify a consistent response level, but varies from experiment to experiment on the basis of assay design. Regulatory limits are therefore set at varied levels of risk dictated by experimental design, not by biologic relevance.

(3) Experiments that have fewer animals tend to result in larger NOAELs, so poor experimental designs are rewarded. When a NOAEL cannot be determined, a LOAEL approach with the addition of a safety factor of 10 is the recommended option. That results in an overconservative calculation.

The Environmental Protection Agency (EPA) Science Advisory Board has challenged the regulatory scientific community to develop improved methods for RfD calculation (U.S. EPA 1991, 1995). Such revisions have been discussed in many specific documents (see annotated bibliography) and policy reviews (for example, National Research Council (NRC), 1994). Development of benchmark dose methods is one approach that has been taken to address the challenge.

The Benchmark Dose

One procedure proposed to replace the use of NOAELs is the use of benchmark dose methods. The benchmark dose (BMD) is usually defined as the lower confidence limit on the dose that produces a specified magnitude of change in a specified adverse response. For example, a BMD_{10} would be the dose at the 95% lower confidence limit on a 10% response, and the benchmark response (BMR) would be 10%. In this example, ED_{10} refers to the dose that produces a 10% excess proportion of abnormal responses. Figure 1 shows this BMD.

The BMD is determined by modeling the dose response curve in the region of the dose response relationship where biologically observable data are feasible. As proposed, the benchmark dose method is not used for extrapolation to low doses where biological

responses can only be estimated.

Both quantal and continuous data can be evaluated with the BMD approach. The most common BMRs are 1%, 5%, and 10% change in response for quantal endpoints. The appropriateness of using these response levels will be discussed in the following sections. A discussion of the appropriate responses to model for continuous endpoints is also given below. Discussions will include an evaluation of various dose response models that have been tested for continuous responses.

Many of the discussions in this paper use a comparison of the traditional NOAEL approach with BMD methods. We do not consider these dose comparisons to be comparisons with a true "gold standard," but rather comparisons with relevant, currently used response indicators.

Applications of Benchmark Dose Methods: Comparisons with Traditional NOAEL Approaches

Developmental and Reproductive Toxicity

Detailed series of investigations of the BMD and its application for noncancer endpoints have been conducted by Allen, Kavlock, Kimmel, and me (Faustman *et al.*, 1994; Allen, 1994a, b; Kavlock *et al.*, 1995). The purposes of the investigations were to assemble an appropriately large and diverse noncancer database to test risk assessment methods, to evaluate and compare the application of traditional NOAEL and generic benchmark methods to this database, to evaluate the impact of incorporating endpoint-specific information in the benchmark modeling process, and to evaluate a series of approaches for the application of BMD methods to both quantal and continuous endpoints.

Faustman *et al.* (1994) describe the developmental toxicity database. It consisted

of 246 experiments (segment II studies) and over 1,825 specific endpoints that were evaluated, which included assessments of visceral and skeletal malformations, growth retardation, and lethality. Reports were obtained from the National Toxicology Program, EPA, and four commercial laboratories (74% of all studies). Characteristics of the database are presented in Tables 2 and 3. About half the experiments evaluated had one or more significant endpoints. The percentage of experiments with significant endpoints increased when fetal body weight changes were included in the analysis. However, Table 4 shows that not all chemicals cause lethality or malformations under the current experimental testing conditions that usually include a high dose with some sign of maternal toxicity.

Unlike most cancer studies, most of these studies had four or five dose groups (175 and 51 experiments, respectively), as shown in Table 2. Table 2 also shows that 39% of the studies had a dose ratio (ratio of highest to lowest dose) of more than 10:1 and that in almost 20% of the cases, adjacent doses differed by a factor of more than 4. Such spacing differences result in large differences in potential NOAELs that can be chosen for these studies with such a traditional approach for RfD determination.

The database was used to compare generic BMD estimates with NOAELs. Two generic BMD models were used for these analyses. A Weibell model (QW) was used for fitting quantal endpoints such as percentage of adversely affected litters. The equation used for this model was

$$P(d) = 1-\exp(-(\alpha+\beta^{*}d^{\gamma}))$$

where the probability of an adverse response at a specific dose is P(d), and α, β, and γ are parameters estimated from the dose response curve (Crump, 1984).

For endpoints for which a continuous measure of response was evaluated (for example, proportion of adversely affected fetuses per litter), a continuous power model was used. It modeled m(d), the mean proportion of adversely affected fetuses in the

group at dose d:

$$m(d) = (\alpha + \beta^{*} d^{\gamma})$$

with the parameters α, β, and γ estimated from the dose response curve (Crump, 1984).

A series of comparisons were made for the quantal NOAELs (QNOAELs) and the quantal BMDs (QBMDs) generated for the same endpoints. Likewise, the continuous NOAELs (CNOAELs) were compared with the continuous BMDs (CBMDs). For all those comparisons, the maximum likelihood estimate (MLE) and BMDs for the lower bound on dose for 10%, 5%, and 1% response rates were evaluated.

For QBMDs, the $QBMD_{10}$ was most equivalent to the QNOAELs. The mean and median ratios of QNOAEL to $QBMD_{10}$ were 2.9 and 2.0, respectively. Thus, for the quantal endpoints, the 95% lower bound estimated on the 10% response ($QBMD_{10}$) was about 2 times lower than the QNOAEL dose (without confidence limits). These observations can be seen in Figure 2.

For CBMDs, the $CBMD_{05}$ was most equivalent to the CNOAELs. The mean and median ratios of CNOAEL to $CBMD_{05}$ were 1.2 and 0.96, respectively. Thus, for these comparisons, the 95% lower bound estimated on the 5% response ($CBMD_{05}$) was about equivalent to the CNOAEL dose (without confidence limits) (See Figure 3).

This study showed that generic BMD models can be used to fit these dose response curves; goodness-of-fit tests rejected fit in only 1% of the quantal examples and 4% of the continuous examples.

Nonconvergence of the models was observed only when nonmonotonic dose response patterns were seen, which resulted from an increase in response rate followed by a decrease in response as doses increased. This type of dose response curve is problematic for all types of evaluations.

This study showed that QNOAELs should not be viewed as "risk free" or "no

adverse effect levels" inasmuch as they were associated with risks of greater than 10% and the CNOAELs were associated with risks of 5%. The risks at the NOAELs were shown to vary, whereas the risks at the BMDs were set. The variation of a NOAEL could have important implications when RfD methods are evaluated. For example, the application of uncertainty values is based on the idea that the NOAEL represents the no observed adverse effect level, not that it sometimes represents the 20% effect level and sometimes the 5% effect level.

The study also evaluated the relationship of LOAELs to NOAELs. Figures 4 and 5 show that if the CLOAEL is divided by 10, as is the recommended regulatory procedure when no NOAEL is determined, the resulting value is 4 to 5 times smaller than the $CBMD_{05}$. The $QNOAEL_{10}$ was about 1.5 times smaller. Thus, the current traditional approaches for using a 10-fold default value for converting a LOAEL to a NOAEL is overconservative.

Allen *et al.* (1994b) evaluated the impact of incorporating endpoint-specific considerations into the statistical models used for BMD estimation. Three additional models were evaluated that included factors to account for nonindependence of observations (for example, observations of fetuses in the same litter) and the impact of other factors, such as litter size. These also evaluated the sensitivity of the models to incorporation of a threshold assumption. The models evaluated included a variation on the Rai and Van Ryzin (1985) model (referred to here as the RVR model), log-logistic model (Kupper *et al.*, 1986), the Log model, and a model developed by Kodell *et al.* (1991) (the NCTR model). Figure 6 shows the equations for these models. Parameter estimates are described in detail in Allen *et al.* (1994b).

As revealed by goodness-of-fit (GOF) statistics, the models that incorporated litter size as a covariable had an improved ability to fit developmental toxicity data. Inclusion of a threshold dose parameter did not seem to affect model fit. Maximum likelihood

limits (MLL) of the log-likelihood were compared across the models to provide an indication of the ability of the models to describe the dose response patterns and variations in litter responses across the curves. These MLLs were compared across the models to provide an indication of the ability of the models to describe the dose response patterns and variations in litter responses across the curves. Significantly higher MLL values were seen for the Log model and these higher MLL values for the Log model were not endpoint-specific. Despite the differences in curve-fitting properties, the BMD_{05} estimates made by the Log, NCTR, and RVR models are very similar (see Allen *et al.*, 1994b for details).

Figures 7 and 8 show that the lower bound estimates at 5% response for the Log model ($LBMD_{05}$) are similar to the $CBMD_{05}$, QNOAEL, and CNOAEL values for these data sets. As was observed with the generic BMD models, the $QBMD_{05}$ is again lower than the $LBMD_{05}$ by a factor of about 3-5.

In summary, the Allen *et al.* paper (1994b) shows that models for noncancer endpoints can incorporate endpoint-specific information. Curve fit is improved, but model predictions are minimally affected. Combined, these three models were able to model all but 45 of the 607 endpoints with a significant dose response curve. There were a few examples of significant endpoints that could be modeled only by a subset of these models. In half of the cases where curve fit was not possible with any of the models, litter-size information was not available; four of the 16 remaining cases were from a single experiment. Readers are referred to Allen *et al.* (1994b), which includes an extensive discussion of litter size as a parameter for developmental toxicity study dose response modeling. As a battery, these models are useful for dose response assessment for noncancer endpoints.

In our observations, the inclusion of a threshold dose parameter did not change model fit from unacceptable to acceptable. Practically, there might be additional reasons

not to include this parameter. First, the models, especially the RVR model, execute more rapidly and converge more often when the threshold parameter is eliminated. Second, elimination of this parameter adds an extra degree of freedom for χ^2 GOF statistics. Its elimination also minimizes confusion with the existence of a biologic threshold for the endpoint under evaluation. It is important to remember that the threshold parameter in these models is based solely on the observed responses in the study. It indicates only the experimental dose below which the best-fitting model no longer predicts an increase in response rate. Despite these points, discussion about the inclusion or elimination of this parameter has resulted in heated philosophical debate (Barnes *et al.*, 1995).

Kavlock *et al.* (1995) examined a variety of approaches for estimating BMDs of continuous endpoints such as fetal weight changes. One of the challenges to investigators evaluating these types of data is to develop clear definitions of biologically significant effects. To develop such a definition, these investigators evaluated multiple approaches for defining a benchmark response. These approaches included a litter-based approach where change in mean fetal weight in a litter in response to treatment was evaluated (used continuous power model) or a fetus-based approach where decreases in individual fetal weights in a litter were compared with preset levels (used log-logistic model). Figure 9 provides a graphical representation of benchmarks developed for litter-based approaches, including evaluation based on percentage change in mean litter weight, change in mean litter weight in relation to variability in control weight, and mean litter weight reduction compared with a control group weight distribution.

The litter-based approaches evaluated a difference of 5% in mean fetal weight, a decrease in mean litter weight of treated litters to the 25th percentile of mean control litter weight, a decrease in mean litter weight by 2 standard errors, and a decrease by 0.5

standard deviation unit. The fetus-based approach examined various options for defining fetal weight changes based on cumulative frequency distributions of control group weight changes. The evaluations included evaluating a 5% added risk of weighing less than the 5th percentile of control weights and a 10% added risk of weighing less than the 10th percentile of control weights.

Those investigations were conducted on a subset of the developmental toxicity database that has been described in Table 3 but for which individual fetal weight values were available (Faustman *et al.*, 1994 and Kavlock *et al.*, 1995). This subset consisted of 173 developmental toxicity studies. Table 4 shows characteristics of the database and reveals that the effects on fetal weight were seen both in the presence and in the absence of other indicators of developmental toxicity.

Figure 10 shows frequency histograms for the ratios of the BMD to values of NOSTAT (statistically derived NOAELs that are determined in these investigations). Faustman *et al.* (1994) had previously shown that these NOSTAT values did indeed represent NOAEL values derived with expert judgment for this database.

The results reveal that, for all the continuous modeling approaches evaluated, the values are very similar, with ratios between 0.5 and 2 and with rare examples of BMDs lower than the NOAELs by a factor of more than 4. The fact that all of the histograms were skewed to the right suggests that BMDs that differed from NOAELs tended to be numerically greater.

Those studies show that continuous endpoints, such as fetal weight, can be adequately modeled with both log-logistic and continuous power dose response models. In addition, they have demonstrated that several approaches provide BMD values that are on the average similar to each other and to NOAEL values.

Application of BMD methods for continuous endpoints was slow because of lack of in-depth analyses of how to evaluate continuous endpoints. Crump (1984), in the

first of the series of BMD papers, discussed some initial approaches for modeling noncancer continuous data and provided examples for four regression models for estimating increased risk of liver fat content in rats exposed to carbon tetrachloride, decreased body weights in hexachlorobutadiene-exposed rats, and decreased thymus weights in TCDD-exposed rats. His studies suggested that the BMD values at 1% extra response were comparable with the no observed effect level (NOEL).

Gaylor and Slikker (1990) evaluated a four-step process to model developmental neurotoxicity using continuous data to describe the risk of adverse neurologic function after exposure to methylene dioxymethamphetamine. They compared the distribution of control and treated rodent neurotransmitter levels to define an adverse change in neurotransmitter levels as equal to 3 standard deviations below the mean levels in control rodents. The proportion of treated fetuses that reached the adverse effect level was then modeled as a function of dose. (Note that in the paper by Kavlock *et al.* (1995) described in detail above, comparisons with the NOAEL levels were comparable at 2 standard error levels from the mean.)

Quantitative BMD models for continuous endpoints for reproductive toxicity have also been evaluated. Pease *et al.* (1991) used BMD methods to model decreases in sperm count after exposure of rabbits to dibromochloropropene. The BMD was defined as the lower 95% confidence limit on a dose that decreased mean sperm count by 10% compared with control rabbit sperm count. They compared the decreases with epidemiologic data on sperm counts in humans and made cross-species comparisons of biologic effect. Their comparisons suggested that a comparable 10% decrease in sperm count in humans would result in a biologically adverse response equivalent to 60x more severe response in humans. This study not only provides an interesting example of the application of continuous BMD methods for reproductive endpoints, but also shows how important cross-species information can be obtained with specific effect level

comparisons, rather than NOAEL comparisons.

Other investigators evaluating continuous BMD approaches include Catalano *et al.* (1994) who have presented combined analysis methods in which fetal death, weight, and malformation changes can be modeled together. The continuous endpoint of fetal weight is evaluated with a definition of adverse effect on weight as changes of 3 standard deviations below the mean of control animal weights. Again, as for Gaylor and Slikker (1990), these comparisons are at 3 standard deviations from the mean versus 2 standard errors from the mean used for comparisons made by Kavlock et al. (1995); thus, greater differences from traditional NOAEL values were observed.

Crump (1995) has developed a new simplified, generally applicable approach to modeling continuous data. It represents an important addition to the available approaches for modeling continuous data.

In summary, continuous modeling approaches are now available for a diverse spectrum of toxic responses, and their application is no longer impeded by lack of comparisons with currently used NOAEL techniques. A good example of this is in the release of the USEPA Risk Assessment Form guidance document on benchmark dose applications for health risk assessment (Crump *et al.*, 1995).

Cancer

The BMD approach has been proposed for assessment of nongenotoxic carcinogenic responses. One application of the BMD method for this type of assessment is for the nongenotoxic carcinogen trichloroethylene, TCE (Haag-Grönlund *et al.*, 1995). Previously, the Institute for Environmental Medicine in Stockholm, Sweden, has used NOAEL-LOAEL approaches for general toxicity and for the evaluation of nongenotoxic carcinogens. In this report, Haag-Grönlund *et al.* (1995) applied a generic quantal dose response model (THRESH) and a generic continuous model (THC) to evaluate 80 sets of TCE data. They confirmed the utility of the benchmark approach for

evaluating dose response relationships and standardizing comparisons across bioassays. Their studies also suggest that the BMD_{10} is in the region of the NOAEL value. They noted that, in cases where the dose response curve had a plateau at high doses, the generic polynomial regression models that they applied failed to fit the experimental data. They are doing further evaluations with more flexible dose response models for benchmark calculation.

In the 46 TCE data sets for which a LOAEL could be determined by Haag-Grönlund *et al.* (1995), only 19 had NOAEL values that were also able to be calculated. Thus, for a large fraction of these data sets, a safety factor of 1,000 would have been applied. For nongenotoxic endpoints, these investigators have been applying a safety factor of 1,000-5,000 for these evaluations. The authors conclude by supporting the philosophic advantages of BMD methods. They also state that their studies support the use of a BMD_{10} approach as a regulatory limit similar to that seen using a NOAEL approach for TCE.

Ecological Toxicity

BMD methods can also be applied for ecotoxicity assessments. For example, many ecotoxicity tests, such as daphnia reproductive tests and oyster larvae tests, already use effect level responses to measure the ecological impact of specific environmental pollutants or mixtures. A.J. Bailer has extended his recent work (Oris and Bailer, 1993; Bailer and Oris 1993, 1994) in evaluating dose response relationships and experimental design in aquatic toxicity testing systems to include a BMD method. In recently submitted work, he has evaluated BMD at 25-50% response. Higher levels of benchmark response were used for these endpoints for several reasons, including the power of the ecologic assay to identify NOAEL (Oris and Bailer, 1993), consideration of background of unexposed control responses, and the biologic significance of observed

impacts.

Other Endpoints

Examples of using the BMD response for other biological effects is given by Slob and Pieters (1995). In their talk at the annual meeting of the Society for Risk Analysis, they presented results for applying the method to assess cholinesterase inhibition in human and animal erythrocytes. They defined a critical effect level based on biological significance and used a BMD_x response to describe this effect. Probability distributions for these critical effects were modeled, and the differences in distributions between animals and humans were used to identify species differences instead of reliance on a safety factor approach. Their paper demonstrated two points: (1) the utility of a BMD approach to evaluate other noncancer endpoints (the presentation identified inherent problems in a NOAEL approach); and (2) that such an approach could be used within probability distribution modeling to eliminate a safety factor approach for evaluating interspecies differences. (The paper also evaluated benchmark approaches for both animal and human studies).

Other papers (for example, Gearhart *et al.*, 1995) illustrate the application of BMD methods to epidemiologic studies of mercury-exposed children. In this application different developmental neurobehavioral endpoints are evaluated, and pharmacokinetic information is incorporated into the benchmark assessments. A BMD-based dose response assessment was conducted from the following evaluations of children exposed to mercury in utero: Clay diagnostic survey, events behavior rating score, Wechsler intelligence scale for children, McCarthy scales of children's abilities (motor, memory, general cognition, etc.), Peabody picture vocabulary tests, tests of language development (TOLD), and Burt word recognition. Those assessments suggest that the most sensitive indicators of developmental effects were found with the TOLD grammar

tests, and the least sensitive were the Peabody picture vocabulary tests.

A benchmark dose was calculated for the most sensitive endpoint using human hair analysis for assessment of the human exposure. A physiologically based pharmacokinetic model for mercury was used to predict fetal brain concentrations from in utero exposures using the hair analysis data. A reference dose was calculated using the benchmark concentrations without safety factors since direct human data from a sensitive population was used. This dose was compared with a NOAEL dose. The BMD-based reference doses were higher by 3-8x when compared to the current NOAEL-based EPA standard. This investigation illustrates how BMD methods can be used for assessment of neurobehavioral effects, for assessment of epidemiologic studies, and how it can be used with PBPK models to eliminate the use of safety factors.

Reviews of Benchmark Dose Methods

Examples of the extensive comments and review that BMD methods have received are the workshops held by the Society of Toxicology (Beck *et al.*, 1993), the National Research Council Committee on Risk Assessment Methodology (Mattison and Sandler, 1994), the Risk Assessment Forum (Crump *et al.*, 1995), and the International Life Sciences Institute (ILSI) (Barnes *et al.*, 1995). The workshop organized by the ILSI risk science institute (RSI) was at the recommendation of EPA. It was financially supported by the Ciba Geigy Corporation, Proctor and Gamble, and the American Industrial Health Council. Because of the international interest in these methods, the World Health Organization (WHO) and the Organization for Economic Cooperation and Development (OECD) have been represented at this and the other meetings. The key objectives of these workshops were:

- To evaluate BMD methods with a case study approach
- To evaluate the impact of BMD methods and current RfD values

- To evaluate which endpoints and effects could be modeled with BMD approaches
- To identify potential barriers that would limit the implementation of BMD methods for calculating the RfD
- To identify potential research needs

The summary statements resulting from these workshops are reflected in the following quote from the Research Council workshop:

> Although a formal consensus was not sought, many participants favored the evolution of quantitative techniques for developmental toxicology risk assessment, including replacement of Lowest Observed Adverse Effect Levels (LOAELs) and No Observed Adverse Effect Levels (NOAELs) with benchmark dose methodology (Mattison and Sandler, 1994).

Information in Crump *et al.* (1995) and the ILSI workshop report (Barnes *et al.*, 1995) probably best defines the current status, identifying decisions before implementation, and identifying research needs. Readers are referred to these documents directly for such information.

Incorporation of Biologically Based Dose Response Modeling Approaches

In the examples of BMD model applications reviewed, minimal attention was given to the use of biologically based models versus generic models for fitting the dose response curves. Where developmental toxicity specific information, such as litter size and nonindependence of fetal events within the litter was included, an improvement in the curve fit was seen (Allen *et al.*, 1994b).

It is anticipated that improvements in the models used for curve fitting will occur and that these improvements will include more biologically based and mechanism specific information. As these improvements are made, there should also be an

improvement in curve fitting and in the narrowing of confidence limits surrounding the BMD response. It is also possible that other mechanistic information could be incorporated into these models, such as information from in vitro or mechanistic investigations, rather than just testing or screening bioassay data (Leroux *et al.*, 1996). If BMD methods were employed, a smooth transition to incorporation of this new information could occur. BMD responses could still be calculated and the new information would most likely narrow the confidence limits and result in a higher BMDx. RfD calculations based only on NOAELs would not be able to accommodate this new information. Improvements in RfD calculations would also occur as the separation of PBPK and BBDR modeling breaks down. One can imagine a synthesis of these two components into a biologically based, pharmacokinetic-consistent dose response model. This model could be used in the BMD method and full incorporation of these types of information would occur. This would not be true with NOAEL approaches.

Long Range Impact of Benchmark Dose Methods

The potential impact of BMD methods can be seen if one first reviews how current NOAEL-derived RfD and RfC values are used.

NOAEL-based RfD and RfC values are used for setting acceptable levels of exposure for all noncancer effects. This encompasses a large group of regulatory statutes, different chemicals, and diverse toxic endpoints (reproductive, developmental, neurologic, ecologic, etc.) predicted for varied environmental media (air, water, soil, etc.).

Besides setting specific acceptable regulatory levels for these diverse situations, reference values are used to determine regulatory actions. For example, margin-of-exposure calculations are often calculated on the basis of a comparison of an

environmental contaminant level with an RfD (or RfC) value. State or federal actions are based on specified factors above a value of 1.0. If the RfD values are based on levels of response that are different because of differences in the effect levels that NOAELs represent (sometimes differing nonrandomly because of peculiarities of experimental design for that specific response), then a reviewer can see that actions might be taken according to detection criteria rather than risk based or response based criteria. For example, for ecotoxicity, NOAELs can be used that represent responses ranging from 31 to 100% rather than true measures of no effect (0% response) (Oris and Bailer, 1993). Likewise, for developmental effects, NOAEL values usually represent responses from 1 to 20% response. Those differences are magnified by the differences in uncertainty factors that are applied to the NOAEL, thus making regulatory actions for noncarcinogens highly heterogeneous yet hiding under the guise of consistent action at a "no-effect" response level. Perhaps the most dramatic problems in reference-dose applications occur in their applications by state governments, USEPA, USDOE, USDOD, and the military services in developing systems for setting priorities among hazardous waste sites for cleanup and environmental restoration. For example, in a USDOD system, hazard functions are calculated for media-specific pathways with a margin-of-exposure approach. Figure 11 shows the hazard function for an example in which contaminant A is a carcinogen and contaminant B is a noncarcinogen. The ratio calculated for A is the maximal concentration of A observed in a given medium and is divided by a standard. The standard for carcinogens is calculated by using the slope of the upper 95% confidence limit on a 10^{-6} response level. The ratio is added to the ratio calculated for contaminant B. For noncarcinogens, the standard used is a reference level based on a NOAEL calculation. Thus, two problems occur. First, the standards for noncarcinogens differ widely from the response levels represented by the NOAEL; these ratios represent different "margins of exposure." Second, when the standards for carcinogens and noncarcinogens represent such different response levels, 10^{-6} versus 10^{-}

[1], then these equations are driven solely by cancer and do not represent any consideration of noncancer responses.

These types of differences are represented in other agency methods that include any type of cross-endpoint comparisons (for example, multi-attribute utility functions in the USDOE Environmental Remediation Priority System (ERPS), which includes human health and ecologic health factors in a single utility function with economic factors). What is needed is a method that allows for a common metric of risk or responses to allow for balanced comparisons for all those varied effects. I propose that BMD methods not only provide a common metric whereby response levels of all noncancer responses can be compared, but also offer a common mathematical metric by which most factors can be compared.

One could imagine that adverse response levels could be defined not only for cancer and noncancer effects and ecological effects, but also for diverse nontoxicity-based responses. For example, DOE has recently defined social, cultural, and economic effects in its risk-based assessment approaches. However, at present, other than purely qualitative approaches, there is no response-based approach to evaluate these effects. Yet, these responses are found in common risk management matrices. One approach to improve consistency in the way such considerations across effects are made would be to define a benchmark response for these very different responses. Thus, a 25% reduction in access to culturally important ceded lands might be defined as a significantly adverse response level by Tribal Nations. In addition, a 15% decrease in jobs could also be an example of an adverse response level that could be defined by city governments and stakeholders affected by remediation actions. (These response levels are only examples; they do not imply specific adverse impact levels that should be used.)

If such a common response-level approach is used, the consistency of risk management decisions should improve, and the ability of risk managers to compare

potential actions and impacts should dramatically improve. Obviously, the approaches just described go beyond the currently evaluated uses of BMDs. However, the Commission on Risk Assessment and Risk Management should fully consider the broader implications of deciding to endorse a BMD and the value-added for these very broad applications.

Case Example of Cost Effectiveness of Benchmark Dose Methods

Advantages of the BMD method is its ability to use experimental data from studies that were not able to ascertain a NOAEL and thus can minimize the need for additional costly experiments (costly financially and in animal lives). This approach also provides an attractive alternative to the calculation of an RfD with a 1,000-fold default assumption on a LOAEL value.

An excellent example of the cost-effective advantages of the BMD methods is given in the case example provided here in its entirety (personal communication from P. Strong). In addition to the advantages listed in the preceding paragraph, this example shows an ability to combine data from multiple studies for metaanalysis. A NOAEL approach would not allow for this type of analysis.

Questions and Answers on Benchmark Issues

a. Are BMD methods proposed as a mechanistic approach to risk assessment?

No.

b. Are BMD methods proposed for extrapolation?

No. One of the key features of BMD methods is looking at dose response

relationships in the range of the dose response curve where biologically observable data is possible. They are not designed for extrapolation beyond this range to very low doses beyond the biological evaluation range. They can, however, be used to identify responses that more closely resemble NOAEL in responses when only LOAELs and not NOAELs are identified for a specific bioassay.

c. Will BMD methods be able to utilize mechanistic or biologically based information?

Yes. One simple example of this is found in the paper by Allen *et al.* (1994b) where incorporation of litter size and nonindependence was included in the models used to evaluate litter effects. This is one example of how biologically based information can be used to improve the dose response models available and thus to improve the biologic basis of curve-fitting. NOAEL values are limited in their ability to be responsive to such mechanistic information.

d. When has the NOAEL-based approach failed?

Frequently NOAEL values are not determined by study designs. The NOAEL approach has then failed, and LOAELs are used as the default. When a LOAEL value is used for determining RfD values, an extra uncertainty factor of 10 is added. Our studies (Allen *et al.*, 1994a, b) have shown that this is too conservative and identifies a regulatory value that is below the 10% response level.

e. Why is there a need to move from a NOAEL-based standard to a BMD based standard?

One dramatic example of the problems inherent in identification of NOAELs was given by H.B.W.M. Koëter (OECD Environment Directorate) at the recent NAS Symposium on New Approaches for Assessing the Etiology and Risks of Developmental Abnormalities from Chemical Exposure (Dec. 1995, Washington, DC; NAS, 1996). Dr. Koëter discussed a study used in the 1994 OECD pesticide project review that

compared the identification of NOAELs for reproductive and developmental toxicity as assessed by regulatory agencies of five OECD member countries. NOAELs were commonly found to differ by a factor of 20 to more than 30 between assessment groups and between countries. Extreme cases differed much more. Variations in terminology accounted for some of the differences, but differences in the interpretation of dose response information and statistical significance also contributed to the large differences. Given the known variability in response levels at the NOAEL from 1 to 20%, and the variation in application of safety or uncertainty factors, the total differences between countries and between experiments could result in combined differences of up to 1,000 fold for setting of acceptable levels of exposure. A similar exercise was undertaken by the National Toxicology Program at the National Institute of Environmental Health Sciences (NIEHS) and similar problems in setting NOAEL levels were noted. BMD methods would address many of these problems.

The analysis presented in Allen *et al.* (1994b) illustrates one of the improvements in the BMD approach over the NOAEL approach; it allows for proper accounting for sample size with statistically appropriate lower confidence limits on dose. This emphasizes the need for models that represent underlying properties of the data. In the absence of such models, the generic models might underestimate the variability and overestimate the BMDs. For developmental toxicity, such considerations have been accomplished. For other responses, consideration of data specifics still needs to be refined to take full advantage of the improvements offered by the BMD approach.

Alternative modeling approaches are under development and include application of generalized estimating equations and considerations of multiple outcomes (Chen *et al.* 1991; Ryan, 1992; Catalano *et al.*, 1994; Krewski and Zhu, 1994, 1995).

Ultimately, such modeling approaches should extend to the development of true biologically based models incorporating, in one approach, toxicokinetic, toxicodynamic

and mechanistic information.

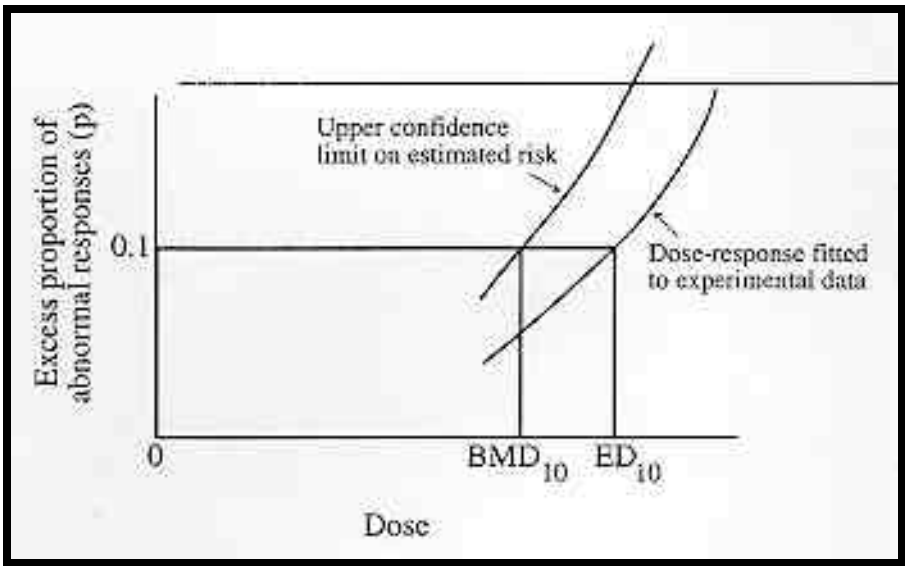

Fig. 1. Benchmark Dose

(Faustman and Omenn, 1995)

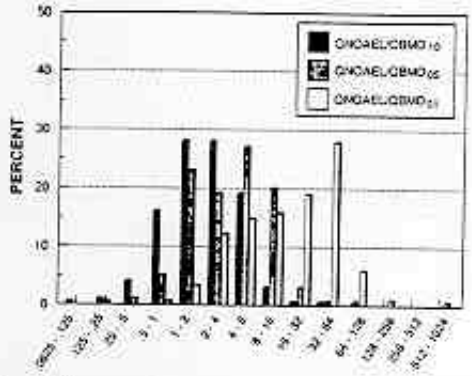

Fig. 2. Histograms of QNOAEL/QBMD ratios. Distributions of the ratios of QNOAELs to QBMDs for the 407 endpoints with significant quantal trends that had QNOAELs and satisfactory fits by the QW model. Factors of 2 define the intervals of ratio values (*x*-axis). The mean (±SD) and median values for $QNOAEL/QBMD_{10}$ were 2.9 (±3.9) and 2.0, respectively; for $QNOAEL/QBMD_{05}$ they were 5.9 (±8.4) and 4.0; for $QNOAEL/QBMD_{01}$ they were 29 (±44) and 19.

Allen *et al.*, 1994

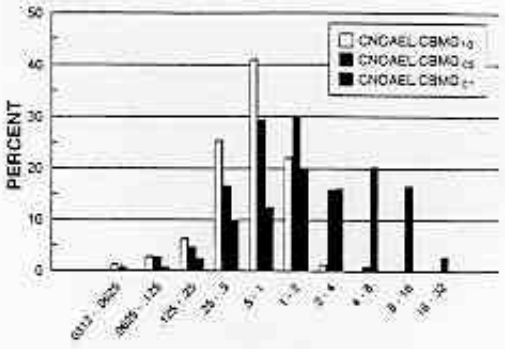

Fig. 3. Histograms of CNOAEL/CBMD ratios. Distributions of the ratios of CNOAELs to CBMDs for the 486 endpoints with significant continuous trends that had CNOAELs and satisfactory fits by the CP model. Factors of 2 define the intervals of ratio values (*x*-axis). The mean (±SD) and median values for $CNOAEL/CBMD_{10}$ were 0.72 (±0.44) and 0.62, respectively; for $CNOAEL/CBMD_{01}$ they were 1.2 (±0.88) and 0.96; for $CNOAEL/CBMD_{05}$ they were 4.3 (±4.5) and 2.5.

Allen *et al.*, 1994

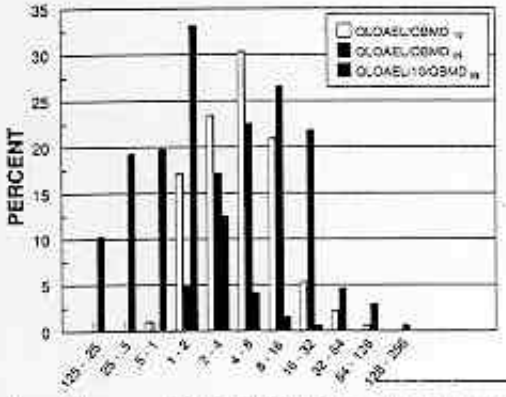

Fig. 4. Histograms of QLOAEL/QBMD ratios. Distributions of the ratios of QLOAELs to QBMDs for the 424 endpoints with significant quantal trends and satisfactory fits by the QW model. Factors of 2 define the intervals of ratio values (*x*-axis). The mean (±SD) and median values for $QLOAEL/QBMD_{10}$ were 7.4 (±9.6) and 4.9, respectively; for $QLOAEL/QBMD_{01}$ they were 15 (±21) and 10; for $QLOAEL/10/QBMD_{10}$ they were 1.5 (±2.1) and 1.0.

Allen *et al.*, 1994

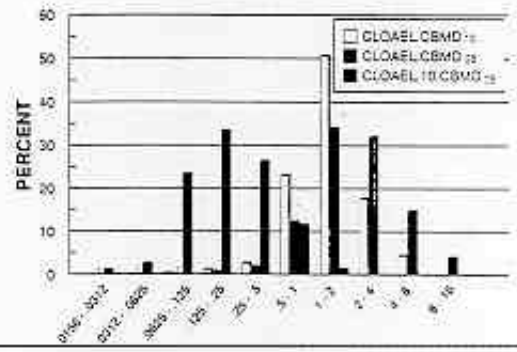

Fig. 5. Histograms of CLOAEL/CBMD ratios. Distributions of the ratios of CLOAELs to CBMDs for the 507 endpoints with significant continuous trends and satisfactory fits by the CP model. Factors of 2 define the intervals of ratio values (*x*-axis). The mean (±SD) and median values for $CLOAEL/CBMD_{10}$ were 1.6 (±0.99) and 1.3, respectively; for $CLOAEL/CBMD_{01}$ they were 2.7 (±2.1) and 2.0; for $CLOAEL/10/CBMD_{10}$ they were 0.27 (±0.21) and 0.20.

Allen *et al.*, 1994

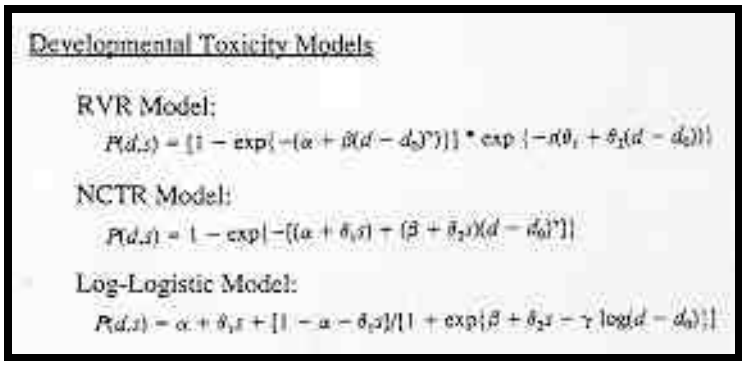

Fig. 6. Models used in the study of BMD approaches for developmental toxicity.

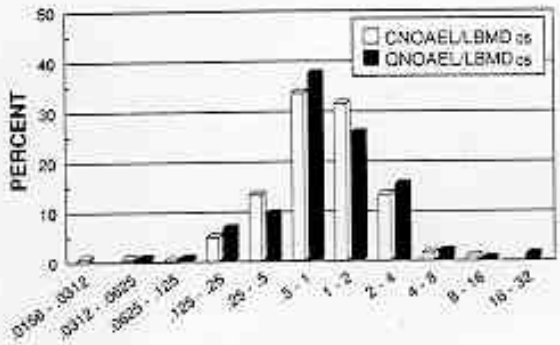

Fig. 7. Representative histograms of $CNOAEL/LBMD_{05}$ and $QNOAEL/LBMD_{05}$ ratios. Ratios are shown for those endpoints with satisfactory fit by the LOG model. The ratio intervals represent factors of 2. The mean (±SD) and median values for $CNOAEL/LBMD_{05}$ were 1.3 (±1.4) and 0.97, respectively; for $QNOAEL/LBMD_{05}$ they were 1.5 (±2.4) and 0.96.

Allen *et al.*, 1994

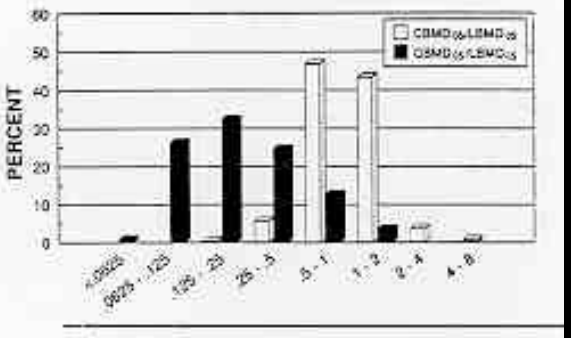

Fig. 8. Representative histograms of $CBMD_{05}/LBMD_{05}$ and $QBMD_{05}/LBMD_{05}$ ratios. Ratios are shown for those endpoints with satisfactory fit by the LOG model. The ratio intervals represent factors of 2. The mean (±SD) and median values for $CBMD_{05}/LBMD_{05}$ were 1.1 (±0.60) and 0.99, respectively; for $QBMD_{05}/LBMD_{05}$ they were 0.30 (±.29) and 0.21.

Allen *et al.*, 1994

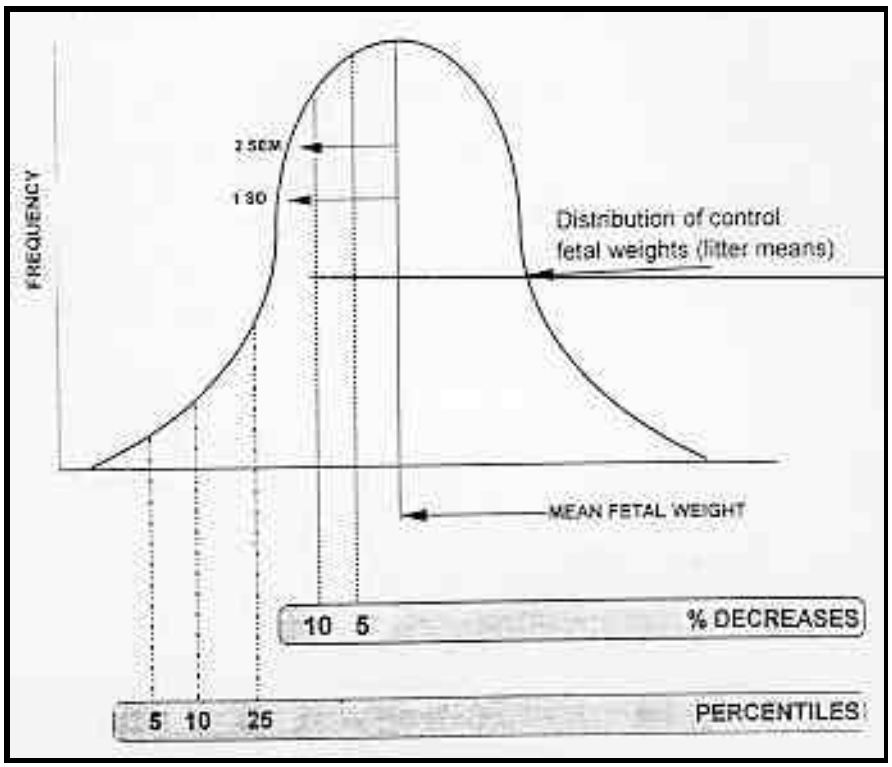

Fig. 9. Graphical representation of the various BMEs used in determination of the litter-based BMDs. The hypothetical distribution of control mean litter weights is depicted, along with indications of the locations on the distribution curve for reductions in mean litter weight of 5 and 10%, reductions in mean litter weight equal to the 5th, 10th, and 25th percentiles, as well as reductions equal to the magnitude of the standard deviation and two standard errors of the mean. The positions of the various BMEs relative to each other have been exaggerated for illustration purposes. Kavlock *et al.*, 1995

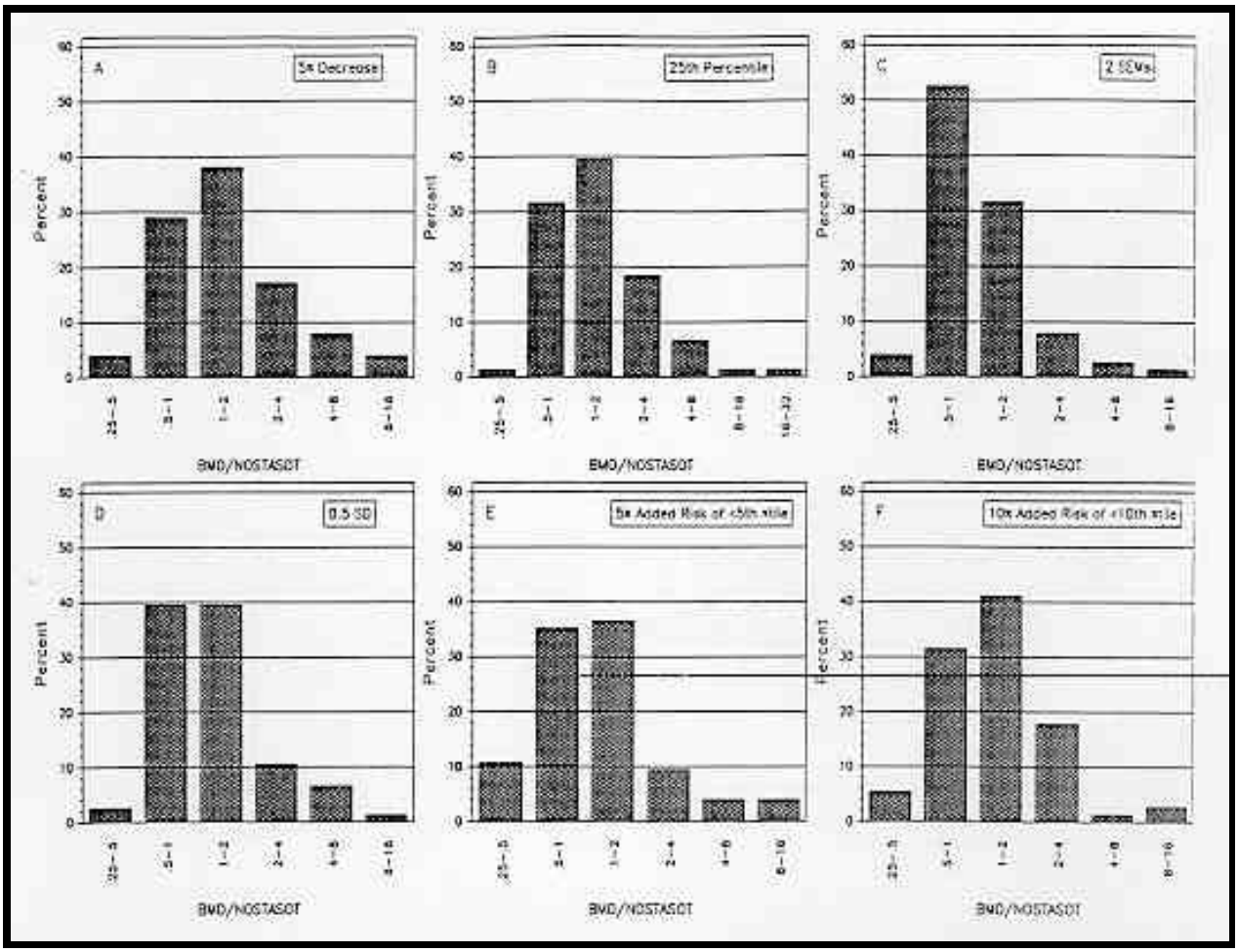

Fig. 10. Frequency histograms of the ratios of the BMD to NOSTASOT for each of the four litter-based and two fetus-based approaches to estimating benchmark doses. Ratios are grouped in categories of a factor of 2. Note that for all approaches, the majority of datasets yield ratios that lie between 0.5 and 2. The distributions tend to be skewed to the right, indicating that when BMDs differed appreciably from the NOSTASOT, they tended to be numerically greater. Note that there were no examples in which the BMD was more than a factor of 4 lower than the NOSTASOT. Kavlock *et al.*, 1995

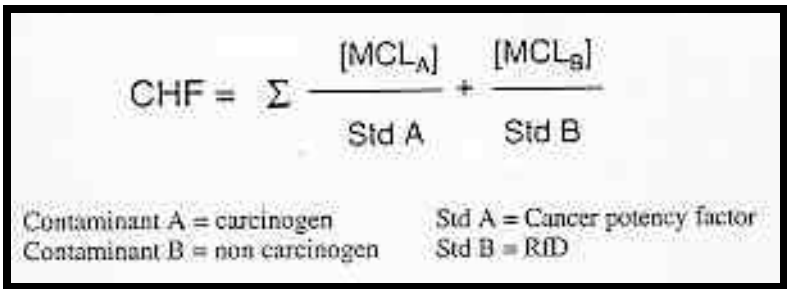

Fig. 11. Site evaluation model contaminant hazard factor (CHF).

Table 1. Incidence rate and estimated economic costs* of cerebral palsy and 17 of the most clinically important birth defects, by condition and type of cost - United States, 1992.

Condition	Incidence rate†	Direct costs		Indirect costs** (millions)	Total costs†† (millions)	Cost per new case (thousands)
		Medical§ (millions)	Nonmedical¶ (millions)			
Nervous system						
Cerebral palsy§§	12.3	$ 852	$445	$1,128	$2,428	$503
Spina bifida	4.2	$ 205	$ 43	$ 241	$ 489	$254
Cardiovascular						
Truncus arteriosus	1.1	$ 108	$ <1	$ 101	$ 210	$505
Single ventricle	1.3	$ 62	$ <1	$ 110	$ 173	$344
Transposition/ Double outlet right ventricle	4.9	$ 188	$ 4	$ 344	$ 515	$267
Tetralogy of fallot	3.5	$ 185	$ 4	$ 171	$ 360	$262
Alimentary tract						
Tracheo-esophageal fistula¶¶	2.9	$ 62	—	$ 103	$ 165	$145
Colorectal atresia	4.5	$ 57	—	$ 162	$ 219	$123
Cleft lip or palate	17.7	$ 97	$ 20	$ 599	$ 697	$101
Atresia/stenosis of small intestine¶¶	3.8	$ 63	—	$ 47	$ 110	$ 75
Genitourinary						
Renal agenesis	4.3	$ 25	—	$ 399	$ 424	$250
Urinary obstruction¶¶	10.4	$ 48	—	$ 297	$ 343	$ 84
Musculoskeletal						
Lower-limb reduction	2.2	$ 17	$ 12	$ 139	$ 187	$199
Upper-limb reduction	4.4	$ 11	$ 24	$ 135	$ 170	$ 99
Abdominal wall						
Omphalocele¶¶	1.9	$ 28	—	$ 104	$ 132	$178
Gastroschisis¶¶	2.8	$ 55	—	$ 54	$ 109	$108
Chromosomal abnormality						
Down syndrome	10.5	$ 279	$389	$1,180	$1,848	$451
Other						
Diaphragmatic hernia¶¶	3.7	$ 63	—	$ 302	$ 364	$250
Total¶¶	**83.8**	**$2,104**	**$887**	**$5,039**	**$8,031**	**$244**

* Costs (in 1992 dollars) are based on lifetime estimates for the 1988 birth cohort in California adjusted for differences in births and costs between California and the nation and for cost inflation during 1988–1992. Future costs are discounted at 5% (9).

† Per 10,000 live births.

§ Medical costs were estimated through the second year of life only for persons born with tracheo-esophageal fistula, atresia/stenosis of small intestine, urinary obstruction, gastroschisis, omphalocele, or diaphragmatic hernia and through age 17 years for those born with colorectal atresia. For all other conditions, medical costs were estimated through age 65 years.

¶ Includes developmental services costs for persons born with cleft lip or palate, spina bifida, Down syndrome, and cerebral palsy, and special education costs for persons born with these conditions as well as for those born with upper- or lower-limb reduction and heart anomalies.

** Includes indirect costs of illness for persons born with cleft lip or palate, spina bifida, Down syndrome, cerebral palsy, upper- or lower-limb reductions, and heart anomalies, and indirect costs resulting from first-year mortality for persons born with any of the conditions except spina bifida, cerebral palsy, and Down syndrome. For the latter three conditions, indirect costs attributable to excess mortality were estimated through ages 9, 17, and 65 years, respectively.

†† Row totals may not equal row sums because of rounding.

§§ Estimates of incidence were based on the proportion of 3-year-olds with cerebral palsy.

¶¶ Column totals are less than column sums because total cost estimates reflect a downward adjustment to avoid duplication when a child had more than one condition.

(MMWR, 1995)

Table 2. Characteristics of the database by number and spacing of dose groups.

No. dose groups[a]	Total studies	Ratio of high dose/low dose[b]						Average spacing between adjacent dose groups[b]				
		1–2	2–3	3–5	5–7	7–10	10+	1–1.5	1.5–2	2–3	3–4	4+
2	2	—	—	—	—	—	—	—	—	—	—	—
3	8	4	1	1	1	1	0	0	4	1	0	3
4	175	2	7	36	13	55	62	2	31	38	68	36
5	51	0	5	6	4	10	26	6	13	18	7	7
6	8	0	0	1	0	0	7	0	2	4	1	1
7	2	0	0	0	0	0	2	0	0	2	0	0
	246	6	13	44	18	66	97	8	50	63	76	47

[a] Includes the control group.
[b] Does not include control group.

Faustman *et al.*, 1994

Table 3. Characterization of database by source, species, and significance of endpoints.

Source	Species	Total number of experiments	Number with: One or more significant endpoints[a]	Number with: No significant endpoints[a]
NTP/EPA	Mouse	21	18	3
	Rabbit	10	8	2
	Rat	32	25	7
	Hamster	1	1	0
Wil Labs	Rabbit	46	20	26
	Rat	73	38	35
Mobil	Rat	12	10	2
Argus	Rabbit	4	3	1
	Rat	5	2	3
IRDC	Rabbit	10	6	4
	Rat	32	10	22
Total		246	141	105

[a] A significant endpoint was one for which either the quantal test or the continuous test for dose-related trend was significant at the 5% level.

Faustman *et al.*, 1994

Table 4. Database evaluation for potential multiple effects on fetal evaluation of endpoints of: prenatal death, malformation and growth retardation.

Dose-related effect on prenatal death or malformations?	Dose-related decrease in fetal weight?		
	Yes	No	Total
Rats			
Yes	33	16	49
No	21	26	47
Total	54	42	96
Rabbits			
Yes	9	15	24
No	2	30	32
Total	11	45	56
Mice			
Yes	16	1	17
No	3	0	3
Total	19	1	20
All species*			
Yes	59	32	91
No	26	56	82
Total	85	88	173*

* Includes one hamster study (positive for both death/malformation and decreased fetal weight) in addition to those in the rat, rabbit, and mouse.

Kavlock *et al.*, 1995

Case Study

BORAX

U.S. Borax Inc.
26877 Tourney Road
Valencia, CA 91355-1847
telephone (805) 287-5400
fax (805) 287-5495

Elaine M. Faustman, Ph.D., DABT
Associate Professor of Toxicology
School of Public Health and Community Medicine
University of Washington
F-561C Health Sciences Center
Seattle, Washington 98195

November 17, 1995

Dear Elaine:

This letter is being submitted in response to your request for information on boric acid that would serve as an example of "Cost Effectiveness" of the benchmark dose (BMD) approach to risk assessment. You were also interested in information related to the combination of data from two similar studies resulting in greater statistical confidence in the calculated BMD. And finally, we discussed the excellent agreement between the NOAEL and the BMD values which were obtained from these studies. Each of these three issues is addressed separately below.

COST EFFECTIVENESS

We believe that the boric acid case provides an example where significant time, cost ($200,000) and animal usage could have been saved if the benchmark dose approach had been fully accepted at the time. Therefore, the following example is provided:

In May of 1990 under the National Toxicology Program, the National Institute of Environmental Health Sciences (NIEHS) issued a final report on the "Developmental Toxicity of Boric Acid in Sprague Dawley Rats". This work was later published in the peer reviewed literature [JJ Heindel et al. *Fund. and Appl. Toxicol.*, 18, 266-277 (1992)]. A NOAEL was not obtained from this rat study. The lowest dose tested was considered a LOAEL, based on a 6% decrease in fetal body weight. (A 5% decrease is generally the limit of statistical significance for this type of study.) Nevertheless, concern was expressed by regulatory agencies, and the ATSDR Minimal Risk Level (MRL) was calculated utilizing a 1000-fold uncertainty factor, based on the LOAEL for decreased fetal body weight, to obtain a value of 0.01 mg B/kg bw/day. This ATSDR value is below environmental levels and could, if enforced, result in deficiencies in boron, considering the essential nutrient requirements for plants and the beneficial nutrient effects being reported for animals and humans. [FH Nielsen, "Facts and Fallacies About Boron", *Nutrition Today*, 27, No. 3, 6-12 (1992) and FH Nielsen, "The Saga of Boron in Food: From a Banished Food Preservative to a Beneficial Nutrient for Humans", *Current Topics in Plant Biochemistry and Physiology*, 10, 274-286 (1991)]

20 MULE TEAM

Because of the importance of having a defined NOAEL for risk assessments, particularly as utilized by regulatory agencies, U.S. Borax contracted a follow-up developmental study in rats exposed to boric acid in their diets with the main goal to obtain a NOAEL for developmental effects. The study was contracted with the same laboratory, using the same study director and other personnel, to the extent possible, who had carried out the initial study for NIEHS. Key personnel from NIEHS, who were involved in the original rat study, were also involved on the Scientific Expert Panel for the follow-up study (i.e. JJ Heindel and BA Schwetz). Experimental conditions of the first study were repeated as closely as possible in the follow-up study, except that three lower doses were utilized in addition to two doses overlapping the lowest doses in the first study, plus the control. The second study evaluated the same end points as the original study and also looked closely at rib effects to better define observations from the first study. Another aspect of the follow-up study was the addition of a parallel group of animals at all dose levels to evaluate postnatal recovery.

The follow-up study confirmed the results of the first study and obtained a clearly defined NOAEL based on reduced fetal body weight, the most sensitive effect. (CJ Price et al. "Developmental Toxicity NOAEL and Postnatal Recovery in Rats Fed Boric Acid During Gestation", submitted for publication in *Fund. Appl. Toxicol.*) While plans were being made for this second study, the BMD methodology came to our attention, and we contracted with Bruce Allen, ICF Kaiser, to calculate the BMD based on the original NIEHS developmental study. If the BMD concept had been fully accepted at the time, the follow-up study would not have been necessary. Considering only the prenatal portion of the study, the cost savings would have been approximately $200,000 and use of a significant number of animals could have been avoided. (The new study involved six dose levels and with 30 rats per group averaging over 15 fetuses per rat, there were 180 adults and over 2000 fetuses to examine.) If the additional postnatal portion were considered, the cost savings would be doubled.

COMBINATION OF DATA FROM TWO SIMILAR STUDIES

Subsequently, Bruce Allen calculated the BMD for the second study. (BC Allen et al. "Benchmark Dose Analysis of Developmental Toxicity in Rats Exposed to Boric Acid" submitted for publication in *Fund. Appl. Toxicol.*) Because this study was purposely carried out under conditions as close to those of the first study as possible, he was able to combine the data from both studies, including eight dose levels, to obtain a high degree of statistical confidence. It is unusual for two studies to be available that were conducted so similarly, and it is only because of the remarkable similarity in experimental design that the study data could be combined, as was done by Bruce Allen.

Based on decreased fetal body weight, the BMD from the combined study compares favorably with the BMD obtained from the original study, before the data at three lower doses were available:

BMD from Original Study	BMD from Combined Studies
56 mg boric acid/kg bw/day	59 mg boric acid/kg bw/day

COMPARISON OF NOAEL AND BMD

Since the follow-up study provided an experimentally determined NOAEL based on the most sensitive effect, reduced fetal body weight, this study provides a comparison with the calculated BMD and the agreement is again very good, adding confidence in the values obtained.

NOAEL	BMD from Original Study
55 mg boric acid/kg bw/day	56 mg boric acid/kg bw/day

In other words, the BMD based on the first study would have predicted a value close to that obtained experimentally in the second study. It would have obviated the need for the second study.

If you have any questions about this information, please call me.

Sincerely,

Philip L. Strong, Ph.D., DABT
Manager, Occupational Health and Product Safety
Phone: 805 287 5634 Fax: 805 287 5542

References

Allen BC, Kavlock RJ, Kimmel CA, Faustman EM. Dose-response assessment for developmental toxicity. II. Comparison of generic benchmark dose estimates with NOAELs. Fundam. Appl. Toxicol., 23:487-495, 1994a.

Allen BC, Kavlock RJ , Kimmel CA, Faustman EM. Dose-response assessment for developmental toxicity: III. Statistical models. Fundam. Appl. Toxicol., 23:496-509, 1994b.

Bailer AJ and Oris JT. Modeling reproductive toxicity in ceriodaphnia tests. Environmental Toxicity and Chemistry, 12:787-791, 1993.

Bailer AJ and Oris JT. Assessing toxicity of pollutants in aquatic systems. Case studies in biometry, New York: John Wiley & Sons, Inc., 1994, pp. 25-39.

Barnes DG, Daston GP, Evans TS, Jarabek AM, Kavlock RJ, Kimmel CA, Park C. Spitzer HL. Benchmark dose workshop: criteria for use of a benchmark dose to estimate a reference dose. Reg. Toxicol. and Pharmacol., Apr. 21:pp. 296-306, 1995.

Beck BB, Conolly RB, Dourson ML, Guth D, Hattis D, Kimmel C, Lewis SC. Symposium overview--Improvements in quantitative noncancer risk assessment. Fundam. Appl. Toxicol., 20:478-509, 1993.

Catalano P, Ryan L, Scharfstein D. Modeling fetal death and malformation in developmental toxicity studies. Risk Analysis, 14:629-37, 1994.

Chen J, Kodell R, Howe R, Gaylor D. Analysis of trinomial responses from reproductive and

developmental toxicology experiments. Biometrics, 47:1049-1058, 1991.

Crump K. A new method for determining allowable daily intakes. Fundam. Appl. Toxicol., 4:854-871, 1984.

Crump K. Calculation of benchmark doses from continuous data. Risk Analysis, 15:79-85, 1995.

Crump K, Allen B, Faustman E, Donison M, Kimmel C, Zenich H. The use of the benchmark dose approach in health risk assessment. Risk Assessment Forum, EPA/630/R-94/007, February 1995.

Faustman EM, Allen BC, Kavlock RJ, Kimmel CA. Dose-response assessment for developmental toxicity. I. Characterization of database and determination of NOAELs. Fundam. Appl. Toxicol., 23:478-486, 1994.

Faustman EM and Omenn GS. Risk Assesment. In *Casarett and Doull's Toxicology: The basic science of poisons*, 5th Edition, C. Klaassen (ed.), New York: McGraw-Hill, Inc., 1995, pp. 75-88.

Gaylor DW. Incidence of developmental defects at the no observed adverse effect level (NOAEL). Reg. Toxicol. and Pharmocol., 15:151-60, 1992.

Gaylor DW and Slikker W. Risk assessment for neurotoxic effects. Neurotoxicology, 11:211-218, 1990.

Gearhart JM, Clewell III HJ, Crump K, Shipp A, Silvers A. Pharmacokinetic dose estimates of mercury in children and dose-response curves of performance tests in a large

epidemiological study. Water, Air, and Soil Pollution, 80:49-58, 1995.

Haag-Grönlund M, Fransson-Steen R, Victorin K. Application of the benchmark method to risk assessment of trichloroethylene. Reg. Toxicol. and Phamacol., 21:261-269, 1995.

Kavlock RJ, Allen BC, Faustman EM, Kimmel CA. Dose-response assessments for developmental toxicity. IV. Benchmark doses for fetal weight changes. Fundam. Appl. Toxicol., 26:211-222, 1995.

Kimmel C and Gaylor D. Issues in qualitative and quantitative risk assessment for developmental toxicology. Risk Analysis, 8:15-20, 1988.

Kodell RL, Howe RB, Chen JJ, Gaylor DW. Mathematical modeling of reproductive and developmental toxic effects for quantitative risk assessment. Risk Analysis, 11:587-590, 1991.

Krewski D and Zhu Y. Applications of multinomial dose-response models in developmental toxicity risk assessment. Risk Analysis, 14:613-27, 1994.

Krewski D and Zhu Y. A simple data transformation for estimating benchmark doses in developmental toxicity experiments. Risk Analysis, 15: 29-39, 1995.

Kupper L, Portier C, Hogan M, Yamamoto E. The impact of litter effects on dose-response modeling in teratology. Biometrics, 42:85-98, 1986.

Leisenring W and Ryan L. Statistical properties of the NOAEL. Regul. Toxicol. Pharmacol., 15:161-171, 1992.

Leroux B, Leisenring W, Moolgavkar S, Faustman EM. A biologically based dose-response model for developmental toxicology. Risk Analysis, 16:449-458, 1996.

Mattison DR and Sandler JD. Summary of the workshop on issues in risk assessment: quantitative methods for developmental toxicology. Risk Analysis, 14:595-604, 1994.

Morbidity and Mortality Weekly Report (MMWR). Years of potential life lost before ages 65 and 85--United States, 1987 and 1988. 39:20-2, 1990.

Morbidity and Mortality Weekly Report (MMWR). Economic costs of birth defects and cerebral palsy - United States, 1992. 44: 694-695, 1995.

National Academy of Science (NAS). Symposium entitled New Approaches for Assessing the Etiology and Risks af Developmental Abnormalities from Chemical Exposure, December 1995, Washington, D.C. Proceedings to be published in Reproductive Toxicology, 1996.

National Research Council. Science and judgment in risk assessment. Washington D.C: National Academy Press, 60-64, 1994.

Oris JT and Bailer AJ. Statistical analysis of the Ceriodaphnia toxicity test: Sample size determination for reproductive effects. Environmental Toxicity and Chemistry, 12:85-90, 1993.

Pease W, Vandenberg J, Hooper K. Comparing alternative approaches to establishing regulatory levels for reproductive toxicants: DBCP as a case study. Environ. Health Perspect., 141-155, 1991.

Rai K and Van Ryzin J. A dose-response model for teratological experiments involving quantal responses. Biometrics, 41:1-9, 1985.

Ryan L. The use of generalized estimating equations for risk assessment in developmental toxicity. Risk Analysis, 12:439-447, 1992.

Slob W and Pieters MN. Probabilistic approach to assess human RfDs and human health risks from toxicological animal studies. Proceedings for the Annual Meeting of the Society for Risk Analysis. Abstract D8.04-A:pp. 60. Waikiki, Hawaii, December 3-6, 1995.

USEPA. Guidelines for developmental toxicity risk assessment. Federal Register 56:63798-63826, 1991.

USEPA. Proposed guidelines for neurotoxicity risk assessment. Federal Register 60:52032-52056, 1995.

Additional Reading List and Annotated Bibliography (for selected references)

Alexeef GV, Lewis DC, Ragle NL. Estimation of potential health effects from acute exposure to hydrogen fluoride using a "benchmark dose" approach. Risk Analysis, 13:63-9, 1993.

Allen BC, Strong PL, Price CJ, Hubbard SA, Daston GP. Benchmark dose analysis of developmental toxicity in rats exposed to boric acid. Submitted to Fundam. Appl. Toxicol., April 1995.

Demonstration of the use of BMDs in calculating reference doses. Existing data from two different studies on a variety of endpoints (fetal weight change, rib malformations, enlargement of lateral ventricles, and gross malformations) were combined and analyzed as one dataset. The authors felt that since these study designs were similar, and were carried out by one laboratory, the combination of data would result in a more accurate BMD. The selection of fetal body weight change was found to result in the lowest BMD, at 59 mg/kg/day. The NOAEL for the original studies was 55 mg/kg/day.

Allen BC, Kavlock RJ, Kimmel CA, Faustman EM. Dose-response assessment for developmental toxicity. II. Comparison of generic benchmark dose estimates with NOAELs. Fundam. Appl. Toxicol., 23:487-495, 1994.

Comparison of statistical NOAELs and BMDs. Using the same set of 246 studies as in part I of this article series (Faustman et al., 1994), the authors derived BMDs for risk levels of 1, 5, and 10%. BMDs were calculated using quantal and continuous treatment of the endpoint of each dataset. Developmental toxicity endpoints such as dead implants and both total and specific fetal malformations were used, and modeled both as quantal Weibull responses and

continuous power responses. Comparisons with the NOAELs reported in the first article revealed that continuous NOAELs were most similar to the continuous BMD_{05} levels, with less than one order of magnitude separation for 98% of the datasets. Quantitative NOAELS, however, were typically greater than their respective quantitative BMD_{10}s, due to both lower maximum likelihood estimates (MLE) and wider confidence bounds.

Allen BC, Kavlock RJ, Kimmel CA, Faustman EM. Dose-response assessment for developmental toxicity: III. Statistical models. Fundam. Appl. Toxicol., 23:496-509, 1994.

Comparison of statistical models used in developmental risk assessment and applied for benchmark dose methods. Generalized RVR, LOG, and NCTR models were applied to the 607 datasets, from study I of the series Faustman et al. (1994) , that showed significant dose-response trends. The generalization of the models was done to allow for intralitter effect correlation, confounding by litter size or other variables, and possible thresholds of effect. While all three models were deemed capable of fitting the data, the LOG model was reported to have superior fit to the datasets studied, partly due to its flexibility in handling dependence on litter size and due to its larger maximum log likelihood values. The adjustment for intralitter correlation, in fact, did not prevent litter size from being a significant covariate in response prediction. Threshold parameters did not appear to significantly add to the models. BMD_{05}s were similar for the three models, as well as for generic toxicity dose-response models of the same data.

Allen B, Kavlock RJ, Kimmel CA, Faustman EM. Comparison of quantitative dose response modeling approaches for evaluating fetal weight changes in segment II developmental toxicity studies. Teratology Society Meeting, Tucson, AZ, June 1993. Teratology, May 1993; 47:394.

This presentation gave a comparison of BMD and NOAEL approaches for toxicity data on fetal weights, using a dataset of 20 NTP studies. Continuous BMDs were determined for the following effect levels: 1) 5 and 10% reduction in average of litter mean weights, 2) reduction in average of litter mean weights to the first quartile, tenth, and twentieth of control values, and to 2 SE below their mean, and 3) .05, .1, 1, and 2 standard deviation reduction from control value of average of litter mean weights. A quantal BMD_{05} and BMD_{10} were calculated using a log-logistic model by choosing cutoffs for quantization at the first quartile, tenth, twentieth, and hundredth of control values. BMD_{05}s were in most cases closer to NOAELs than BMD_{10}s. For evaluations of absolute change from control values, the quartile and 2 SE reduction levels proved the most similar to NOAELs. For the third continuous BME above, use of a 1 standard deviation decrease resulted in BMDs most similar to NOAELs. For the log-logistic model, the BMD_{10}s at the 5th percentile cutoff were closest to the NOAELs.

Allen BC, Van Landingham C, Howe RB, Kavlock RJ, Kimmel CA, Faustman EM. Dose-response modeling for developmental toxicity. Society of Toxicology. Seattle, WA, February 1992. The Toxicologist 1992; 12:300.

Allen BC, Van Landingham C, Howe RB, Kavlock RJ, Kimmel CA, Faustman EM. Comparison of approaches to risk assessment for developmental toxicants. Proceedings of the Society for Risk Analysis Annual Meetings, Baltimore, MD, December 1991.

Demonstration of BMD approach, using over 1500 datasets for quantal developmental toxicity responses. Four models were compared for differences in BMD, and were found to be similar for any given effect level. For the most part, the developmental toxicity models fit the data equally well. BMD_{05}s from a log-logistic and the NCTR model were similar to NOAELs, but those from an adapted RVR model and a generic Crump benchmark model tended to be smaller than their respective NOAELs.

Auton TR. Calculation of benchmark doses from teratology data. Reg. Toxicol. and Pharmacol., 19(2): 152-67, 1994.

Bailer AJ and Oris JT. Modeling reproductive toxicity in Ceriodaphnia tests. Environ. Toxicol. Chem., 12:787-791, 1993.

Development of a dose-response model for reproductive toxicity in Ceriodaphnia. Litter size is assumed to be Poisson-distributed and mean litter size is modeled as an exponential function of test concentration. Calculation of confidence intervals and toxicant potency is demonstrated.

Barnes DG, Daston CP, Evans JS, Jarabek AM, Kavlock RJ, Kimmel CA, Park C, Spitzer HL. Benchmark dose workshop: Criteria for use of a benchmark dose to estimate a reference dose. Reg. Toxicol. and Pharmacol., 21:296-306, 1995.

Recommendations from workshop participants on the use of BMDs in calculating RfDs and RfCs. The derivation of BMDs from quantal noncancer data was endorsed, but most participants felt current NOAEL-based standards to be "sufficiently protective," and requiring revision only as appropriate data become available.

Bosch RJ, Wypij D, Ryan LM. A semiparametric procedure for risk assessment with quantitative responses. Teratology Society Meeting, Newport Beach, CA, Teratology, 1995, 51:68.

Crump K, Calculation of benchmark doses from continuous data. Risk Analysis, 15:79-85, 1995.

Explanation of a method for calculating BMDs for continuous health effects. The

method requires the selection of an abnormal response level, but dichotomization of individual responses is avoided through statistical treatment of distributed response at each dose. Abnormality can be defined at the individual or population level. The method is applicable to any dose-response model, for any specified abnormal response rate.

Crump KS, Allen BC, Faustman EM. The use of the benchmark dose (BMD) approach in health risk assessment. Technical Report prepared for EPA (ERG Contract No. 68-C8-0036), Risk Assessment Forum, USEPA, February, 1995 (EPA/630/R-94/007).

Overview of the BMD approach to noncancer risk assessment, including comparison of the NOAEL and BMD methods. This document was developed as a background report intended to generate EPA discussion of the BMD technique. Guidance on response selection and model choice is given, and the use of the BMD method for polynomial, power, Weibull, and log-normal models is demonstrated. The implications of biological assumptions, like the exclusion of a threshold, and statistical assumptions, such as normally or binomially distributed response, are also discussed.

Crump K and Allen B. Prepared for U.S. Department of Labor. Occupational Safety and Health Administration, Washington, D.C., 1988.

Faustman EM, Ponce RA, Seeley MR, Whittaker SG. Experimental approaches to evaluate mechanisms of developmental toxicity. *CRC Handbook of Developmental Toxicity*, CRC Press, Inc., 13-56, 1996.

Faustman EM, Allen BC, Kavlock RJ, Kimmel CA. Dose-response assessment for developmental toxicity. I. Characterization of database and determination of NOAELs. Fundam. Appl. Toxicol., 23:478-486, 1994.

Characterization of NOAEL-based approaches, using data from 246 studies of developmental toxicity. NOAELs were calculated by both continuous and quantal treatment of each dataset. Trend tests indicated that 386 of the 1825 datasets showed a significant trend for both continuous and quantal treatment, 177 datasets had only a significant continuous trend, and 44 had only a significant quantal trend. For those datasets with any significant trend test, 98% had the two NOAELs within one dose level, and 99% had NOAELs separated by two or less dose levels. Additionally, 20 National Toxicology Program studies were used to compare statistical NOAELs with "expert-derived" NOAELs. 92% of these 360 datasets had an identical pair of NOAELs, and over 98% had the two NOAELs within one dose level.

Faustman EM, Allen BC, Van Landingham C, Howe RB, Kavlock RJ, Kimmel CA. Quantitative risk assessment approaches for developmental toxicity: Alternatives to the NOAEL values. VI International Congress of Toxicology, Rome, Italy, July, 1992.

Faustman EM, Kimmel C, Wellington D. Characterization of a developmental toxicity dose-response model. Environ. Health Perspect., 79:229-241, 1989.

Gaylor DW, Slikker W. Risk assessment for neurotoxic effects. Neurotoxicology, 11:211, 1990.

Demonstration of quantal dose-response modeling of neurochemical, neurohistological, and behavioral effects from exposure to methylenedioxymethamphetamine (MDMA). Data are taken from rat and monkey studies are fit to a curve predicting risks to humans. Biomarkers and abnormal levels can be substituted, respectively, for direct effect measurements and adverse effect levels, when the latter information is not available. The authors suggest that dose-response modeling is superior to the use of NOAELs due to increased versatility and fuller use of the available data.

Gearhart JM, Clewell III HJ, Crump K, Shipp A, Silvers A. Pharmacokinetic dose estimates of mercury in children and dose-response curves of performance tests in a large epidemiological study. Water, Air and Soil Pollution, 80:49-58, 1995.

This paper illustrates the application of benchmark methodologies to human epidemiology studies and to neuro-behavioral endpoints. The paper examines the sensitivity of over nine of these endoints in relationship to their ability to detect effects in children following in utero exposure to methyl mercury. The most sensitive indicator for methyl mercury effects was the Test of Language Development grammar tests and the least sensitive endpoint was the Peabody picture vocabulary tests. These authors also included a PBPK model for methyl mercury and illustrate the incorporation of this information into the calculation of a RfD dose that is a factor of three to eight above the current USEPA RfD.

Glowa J. Dose-effect approaches to risk assessment. Neurosci. Biobehav. Rev., 15:153-158, 1991.

Comparison of risk estimates using two different BMD methods. Two datasets (n=10 and n=40) were taken from a published experiment on behavioral effects from acute exposure to toluene in male rats. In both methods, effects were linearly regressed against the natural logarithm of exposure concentration, and a BMD_{10} was calculated. In the first method, dose-response was plotted for each individual rat, and BMDs were accumulated and statistically described. The second method utilized pooled data from all of the rats, so that a single BMD was calculated from each dataset. Results from the methods were equivalent for low risks derived from data with a large sample size, but divergent when higher risks were modeled or smaller sample sizes were used.

Haag-Grönlund M, Fransson-Steen R, Victorin K. Application of the benchmark method to risk assessment of trichloroethylene. Reg. Toxicol. and Phamacol. 21:261-269, 1995.

Evaluation and application of the benchmark method, using trichloroethene (TCE) noncancer toxicity studies. For the eighty datasets used, all BMD_1s and 42% of the BMD_{10}s fell lower than the NOAELs. In addition, 93% of the BMD_{10}s fell lower than their respective LOAELs. Despite these differences, a regulatory guideline exposure value for TCE calculated from the benchmark method gives similar results to those generated by the traditional method. While the fit of the datasets to the polynomial regression models used here was very poor, the authors noted the increased detail and enhanced comparability gained by using the benchmark method.

Jarabek AM, Menache MG, Overton JH Jr, Dourson ML, Miller FJ. The US Environmental Protection Agency's inhalation RfD methodology: risk assessment for air toxics. Toxicol. Ind. Health, 6:279-301, 1990.

Kavlock RJ, Allen BC, Faustman EM, Kimmel CA. Dose-response assessments for developmental toxicity. IV. Benchmark doses for fetal weight changes. Fundam. Appl. Toxicol., 26:211-222, 1995.

Continuation of a series of articles evaluating a large developmental toxicity dataset. (Faustman et al., 1994). Eighty-five of the experiments which contained individual fetal weight data were chosen for this analysis of methods for calculating BMDs from continuous data. A continuous power model was used to model mean response at the litter level, and a log-logistic model was used to model quantized individual response. For the power model, four BMEs were used: 5% reduction in mean fetal weight, a two standard error decrease in mean weight, a half standard deviation decrease in mean weight, and a decrease to the lowest quartile of control

litter mean weights. BMEs for the log-logistic model were evaluated at a 5% additional risk of falling below the 5th percentile of control fetal weight, and a 10% additional risk of falling below the 10th percentile. Strong similarities were reported for all of these BMDs when calculated for any one dataset. BMD values were also comparable to statistically derived NOAELs from the first study.

Kimmel CA, Kavlock RJ, Allen BC, Faustman EM. Application of benchmark dose methodology to data from prenatal developmental toxicity studies. Tox. Letters, 82/83: 549-554, 1995.

Kimmel CA, Ross DC and Francis ES, eds. Proceedings of the workshop on the qualitative and quantitative comparability of human and animal developmental neurotoxicity. Neurotoxicol. Teratology, 12 (3):173-292, 1990.

Kimmel CA, Wellington DC, Farland W, Ross P, Manson JM, Chernoff N, Young JF, Selevan SG, Kaplan N, Chen C, Chitlik LD, Siegel-Scott CL, Valaras G, and Wells S. Overview of a workshop on quantitative models for developmental toxicity risk assessment. Environ. Health Perspect., 79:209-215, 1989.

Kodell R, Howe R, Chen J, Gaylor D. Mathematical modeling of reproductive and developmental toxic effects for quantitative risk assessment. Unpublished report. 1988.

Leisenring WM, Leroux BG, Moolgavkar SH, Ponce RA, Faustman EM. Evaluation of cellular kinetics in biologically based dose-response modeling for developmental toxicology. Teratology Society Meeting, Tucson, AZ, June 1993, Teratology, 47:426, 1993.

Leisenring WM, Leroux BG, Moolgavkar SH, Faustman EM. A biologically based dose-response model for the developmental toxicity of methylmercury. Society of Toxicology Annual

Meeting, Dallas, TX, March 1994. The Toxicologist; 14:39, 1994.

Biologically based mathematical description of the kinetics of the organogenesis process. Using known information on cell kinetics and branching processes, the authors developed a model to describe the timing of such processes as differentiation, migration, growth, and replication. Variability of process kinetics within populations is included, as is the possibility of a threshold of effect. The goal of study was to develop a working model of the effects of methylmercury dose and exposure time on malformation rates, and eventually extend the results to other teratogens. Methylmercury was chosen for the study due to availability of data from numerous in vivo *and* in vitro *studies.*

Malsch P, Proctor D, Finley B. Estimation of a chromium inhalation reference concentration using the benchmark dose method: a case study. Reg. Toxicol. Pharmacol. 20:58-82, 1994.

Nair RS, Stevens MS, Martens MA, Ekuta J. Comparison of BMD with NOAEL and LOAEL values derived from subchronic toxicity studies. Arch. Toxicol. Suppl., 17:44-54, 1995.

This paper reports on a benchmark dose evaluation conducted by Monsanto Company US and Europe jointly with the University of Mississippi, Dept. of Pharmacology to evaluate the potential application of benchmark doses for subchronic toxicity studies. In addition to extending this methodology to other noncancer endpoints, this paper also evaluated the relationship of the maximum likelihood estimate (MLE) for the benchmark dose as well as the BMD 01, 05 and 10 response levels to NOAEL and LOAEL values. Phase II of these studies evaluated the impact of the number of dose levels on the BMD values. These authors concluded that: (1) the BMD approach awarded datasets with good dose-response information (as judged by decreased variability between the MLD and benchmark dose estimates); (2) this study

supported the case of a BMD_{05} - BMD_{10} as having comparable results with NOAEL values from these same studies; (3) the results for continuous studies were more variable and NOAEL values compared to BMD£10; (4) BMD calculations were sensitive to removal of dose response data; (5) BMD approach was consistent with basic principles of toxicology.

Oris JT and Bailer AJ. Statistical analysis of the Ceriodaphnia toxicity test: sample size determination for reproductive effects. Environmental Toxicology and Chemistry, 12:85, 1993.

Calculation of minimum sample sizes needed to detect decreased survivorship and reproductive effects in Ceriodaphnia dubia *toxicity tests. The analysis uses several variance models and a wide range of false positive and false negative error rates. It was found that with the typically used sample size of 10 organisms, fecundity decreases of 31 to 100% , relative to controls, would be necessary in order to meet statistical significance. Since the current test protocol may result in a test of insufficient power, the authors recommend changing the test design so that consideration of detection limits of reproductive inhibition is included in the choice of sample size.*

Rodier PM. Developmental neurotoxicology. Toxicologic Pathology, 18:89, 1990.

Description of differences between developmental and non-developmental neurotoxic injuries. While some developmental injuries are detectable by traditional morphologic evaluation, many effects result from disturbed developmental processes, rather than tissue destruction. These effects, such as misplaced and misoriented neurons, and decreased tissue volume, rarely or never result from adult injury. Accurate evaluation of developmental injury requires familiarity with the temporal scale of functional and structural teratogenic effects.

Rogers JM, Mole ML, Chernoff N, Barbee BD, Turner CI, Logsdon TR, Kavlock RJ. The developmental toxicity of inhaled methanol in the CD-1 mouse, with quantitative dose-response modeling for estimation of benchmark doses. Teratology, 47:175-88, 1993.

Slob W, Pieters MN. Probabilistic approach to assess human RfDs and human health risks from toxicological animal studies. Proceedings for the Annual Meeting of the Society for Risk Analysis Abstract D8.04-A, Pg. 60. Waikiki, Hawaii, December 3-6, 1995.

The authors identify two major drawbacks of using a NOAEL and safety facter approach to derive human RfDs. Uncertainty in the value of the NOAEL, which the authors suggest might be a rather poor estimate of the real no-adverse-effect-level (NOAEL) in the animal, is completely ignored in the standard RFD approach. Multiplication of safety factors for the RFD calculation implies that worst-case assumptions are in fact piled up, making RfD calculations overly conservative. Slob and Pieters propose an alternative approach in which uncertainties are taken into account in a probabilistic fashion. Although the authors present approaches to take into account some of the uncertainties others are (as yet) difficult to assess (e.g. the uncertainty in the interspecies extrapolation factor). In these cases the authors temporarily defaulted to an "educated guess". Advantages of this approach is that it offers the possibility to continuously improve and refine the routine assessment of RfDs, as new knowledge and data become available. The method was also used to assess human health risks together with an uncertainty distribution, given a specified exposure level.

Stanton ME and Spear LP. Workshop on the qualitative and quantitative comparability of human and animal developmental neurotoxicity, work group I report: Comparability of measures of developmental neurotoxicity in humans and laboratory animals. Neurotox. Teratol., 12:261, 1990.

Review of neurotoxicological assessment measures and evaluation of cross-species efficacy. Compounds reviewed include anticonvulsant drugs, ethanol, methylmercury, lead, PCBs, and ionizing radiation, and categories of effect included sensory, motivational, cognitive and motor function, and social behavior. This workshop report identifies good agreement across species within each category, especially with high exposures. Additional conclusions include the following: The EPA test battery correctly identified hazards to humans, but sometimes underestimates risk; all effect categories should be included in assessments; since most neurotoxic effects are unattributable to maternal toxicity, neurotoxic tests should use a maximum upper dose level equivalent to the threshold for maternal toxicity; postnatal maternal exposure results in methodological difficulties; and animal studies should emphasize evaluation during development.

Starr TB. Concerns with the benchmark dose concept. Toxicology Summer Forum, Aspen, CO, July 14, 1995.

Criticism of the use of BMDs to replace NOAELs and LOAELs. The author notes three disadvantages to using the BMD approach in the regulatory setting. The first is the possibility of undue influence from data points at high doses, since the dose-response models in use are not mechanistic. The author specifically highlights problems with using a Weibull model. The second criticism is that the BMD approach is unnecessarily conservative, since it typically yields lower reference doses than the NOAEL approach. Finally, the author states that due to small sample sizes and the limited number of dosing levels, nearly all modeling of available toxicity data results in linearization and BMDs tend to ignore possible thresholds of effect. The author does support the use of BMDs in interpolating between the NOAEL and LOAEL values.

Starr TB. The benchmark dose concept: Questionable utility for risk assessment? Proceedings of the Annual Meeting of the Society for Risk Analysis. Abstract H2.03, pg. 86-87.

Waikiki, Hawaii, Dec. 3-6, 1995.

The author discusses the proposal to use benchmark doses as preferred methodology to the use of No Observed Adverse Effect Levels (NOAELS) in setting "safe" levels for noncarcinogens. The authors describe the benchmark dose method as using the Weibull model to extrapolate below the observable response range to "benchmark" doses that are predicted to yield a specific level of response such as 1% or 5%. Lower bounds are calculated for the benchmark dose at the lower 95% confidence limit. The author raises several concerns about this approach. First, he notes that the Weibull approach is generic, empirical, and lacking a mechanistic foundation, thus, without explicit incorporation of information regarding mechanisms of toxic action, it is "curve-fitting." Second, he notes that given the comparatively small group sizes employed in most toxicology studies a substantial degree of conservatism can be present thus, making the lower bounds relative to central estimates very large. Third, the author states that Crump has also noted that the small number of dose groups in most toxicology studies and their similarly small size nearly always preclude rejection of a linear lower bound. The author suggests that the BMD approach will linearize the regulation of non-cancer endpoints and questions whether we should change our current NOAEL approach. The author raises the question of whether there is any justification for abandoning the traditional threshold concept of toxicology in favor of biologically implausible one hit models for non cancer endpoints? These concerns merit serious consideration by the risk assessment community.

USEPA. Review of statistical methods in carcinogenic dose-response models and evaluation of their appropriateness for modeling dose-response in developmental toxicity. Technical Report, Contract # 5W-6906-NASA, November, 1985.

USEPA. Characterization of a developmental toxicity dose-response model. Technical Report, Contract # 6W-1014-NASA, April, 1986.

Weller EA, Catalano PJ, and Ryan LM. A class of dose-response models for benchmark dose estimation in developmental toxicity studies. Teratology Society Meeting, Newport Beach, CA, Teratology, 1995, 51:69.