DRAFT
DO NOT CITE OR QUOTE

EPA/600/P-96/002A
August 9, 1996
External Review Draft

Benchmark Dose Technical Guidance Document

NOTICE

THIS DOCUMENT IS A PRELIMINARY DRAFT. It has not been formally released by the U.S. Environmental Protection Agency and should not at this stage be construed to represent Agency policy. It is being circulated for comment on its technical accuracy and policy implications.

Risk Assessment Forum
U.S. Environmental Protection Agency
Washington, DC 20460

DISCLAIMER

This document is an external draft for review purposes only and does not constitute U.S. Environmental Protection Agency policy. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

CONTENTS

Figures vi
Preface vii
Authors and Reviewers ix
I. EXECUTIVE SUMMARY 1
A. Introduction 1
B. Decision Process for Carrying Out a Benchmark Dose/Concentration Analysis 3
1. Data array analysis - endpoint selection 3
2. Minimum data set for calculating a BMD/C 4
3. Criteria for selecting the benchmark response (BMR) 4
4. Order of model application 5
5. Determining the model structure 6
6. Confidence limit calculation 7
7. Selection of "equivalent" models 7
8. Selecting the BMD/C 8
9. Use in Risk Assessment 9
II. INTRODUCTION 10
A. Background 10
B. Purpose of This Guidance Document 16
C. Definition of the Benchmark Dose or Concentration 17
III. BENCHMARK DOSE GUIDANCE 18
A. Data Array Analysis - Endpoint Selection 18
1. Selection of Endpoints to be Modeled 19
2. Minimum Data Set for Calculating a BMD/C 19
3. Combining Data for a BMD/C Calculation 20
B. Criteria for Selecting the Benchmark Response Level (BMR) 20
C. Mathematical Modeling 22
1. Introduction 22
2. Order of Model Application 23

3. Determining the Model Structure 23
4. Confidence Limit Calculation 25
5. Selection of "Equivalent" Models 25
6. Selecting the BMD/C 26
IV. USING THE BMD/C IN NONCANCER DOSE-RESPONSE ANALYSIS 28
A. Introduction 28
B. Effect of BMD/C Approach on Use of Uncertainty Factors 28
C. Dose-Response Characterization 29
V. USE OF BENCHMARK-STYLE APPROACHES IN CANCER RISK ASSESSMENT . . 31
A. Use in Hazard Ranking 31
1. Guidance for Developing Comparable ED_{10}s 31
2. Sources of TD_{50}s and ED_{10}s 32
B. Use in Low-Dose Extrapolation 33
1. Proposal for Low-Dose Extrapolation under the 1996 Proposed Guidelines for Carcinogen Risk Assessment 33
2. Issues in Choosing a Point of Departure 37
C. Dose-Response Characterization 38
VI. FUTURE PLANS 40
A. Updating of Guidance Document 40
B. Potential for Use of the BMD/C in Cost-Benefit Analyses 40
1. Background 40
2. Benefits estimates using BMD/C and RfD/C 41
3. Issues in quantifying noncancer health benefits 42
C. Research and Implementation Needs 43
VII. REFERENCES 46

APPENDICES

A. Aspects of Design, Data Reporting, and Route Extrapolation Relevant to BMD/C Analysis 55

1. Design 55
2. Aspects of Data Reporting 55
3. Route Extrapolation 57
B. Selecting the Benchmark Response (BMR) Level 58
1. Biologically Significant Change for Specifying the BMR 58
2. Limit of Detection for Specifying the BMR 59
3. Examples 62
4. Selecting the Critical Power Level for the Limit of Detectability 65
C. Mathematical Modeling 66
1. Introduction 66
2. Model Selection 66
a. Type of endpoint 67
b. Experimental design 68
c. Constraints and variables 69
3. Model Fitting 69
4. Assessing How Well the Model Describes the Data 71
5. Comparing Models 72
6. Using Confidence Limits to Get a BMD/C 73
D. Examples of BMD/C Analyses 76
Example #1: Carbon Disulfide 76
Example #2: 1,1,1,2-Tetrafluoroethane (HFC-134a) 80
Example #3: Boron 85
Example #4: 1,3-Butadiene 91
E. List of Models Planned to be Included in the First Release of the EPA Software 94

LIST OF FIGURES

1. Example of calculation of a BMD 13
2. Dose response curves for models incorporating different forms of spontaneous background response 61
3. Detectable extra risk or additional risk for a sample size of 25 (N) 64
4. Carbon disulfide modeling 79
5. HFC-134a modeling 84
6. Boron modeling - continuous data 89
7. Boron modeling - dichotomous data 90
8. 1,3-Butadiene modeling 93

PREFACE

This draft document on guidance for application of the benchmark dose/concentration (BMD/C) approach in cancer and noncancer dose-response assessment is being developed by a Technical Panel of the Risk Assessment Forum for use by the U.S. Environmental Protection Agency. This document is intended to be used together with the background document on use of the benchmark dose approach published earlier by the Agency (EPA, 1995c). While the major audience for this document is risk assessors within the Agency, it is expected that this guidance will be considered and possibly used by other organizations as well.

This draft guidance document is currently under development, but is being made available at this stage for a peer consultation workshop to be held September 10-11, 1996, at the Holiday Inn, Bethesda, MD. Several experts in the areas of toxicology, statistics, and mathematical modeling have been asked to review the document and provide input at this early stage of development on several issues for which there is ongoing discussion within the Agency. These issues include: the appropriate selection of studies and responses for BMD/C analysis, use of biological significance or limit of detection for selection of the benchmark response (BMR), model selection and fitting, use of the lower confidence limit as the BMD/C, and the default decision approach proposed in this document for the BMD/C analysis.

Our overall goal in developing this document is to have a procedure that is usable, that has reasonable criteria and defaults to avoid proliferation of analyses and model shopping, and that promotes consistency among analyses. Ultimately, we are trying to move cancer and noncancer assessments closer together, using precursor and mode of action data to extend and inform our understanding of risk in the range of extrapolation. We would like to have in one package something that is usable for cancer and noncancer assessments when endpoints are relevant to both.

We also are asking reviewers to comment on how understandable the document is for the general toxicologist and risk assessor. In an effort to achieve a readable and usable document, the primary information on approaches and defaults is contained in the main body of the document, with more detailed discussion of various steps in the process included in the appendices. Several

examples of application of the BMD/C approach using the procedures recommended in this guidance are included in an appendix. In a separate but parallel effort, the EPA is developing user-friendly software for BMD/C analysis to be distributed widely for use by the risk assessment community, and our goal is to make the software consistent with the guidance and default procedures developed in this document.

AUTHORS AND REVIEWERS

TECHNICAL PANEL

Carole A. Kimmel, Chair*

R. Woodrow Setzer, Jr.*

Dan Guth*

Elizabeth Margosches*

Suzanne Giannini-Spohn*

Linda Teuschler*

Jim Cogliano*

Annie Jarabek*

Jeanette Wiltse

Robert MacPhail

Yogendra Patel

William Sette

Carole Braverman

Rick Hertzberg

John Vandenberg

Hugh Pettigrew

*Authors

I. EXECUTIVE SUMMARY

A. Introduction

The US EPA conducts risk assessments for an array of noncancer health effects as well as for cancer. The process of risk assessment, based on the National Research Council (NRC) paradigm (1983), has several steps: hazard characterization, dose-response analysis, exposure assessment, and risk characterization. Risk assessment begins with a thorough evaluation of all the available data to identify and characterize potential health hazards. The next stage of the process, dose-response analysis, involves an analysis of the relationship between exposure to the chemical and health-related outcomes, and historically has been done very differently for cancer and noncancer health effects. The common practice under the 1986 Guidelines for Carcinogen Risk Assessment (EPA, 1986a) was to assume low dose linearity and model tumor incidence data by applying the linearized multistage (LMS) procedure, which extrapolates risk as the 95% upper confidence limit. The standard practice for the dose-response analysis of noncancer health effects has been to determine a lowest-observed-adverse-effect-level (LOAEL) and a no-observed-adverse-effect-level (NOAEL). The LOAEL is the lowest dose for a given chemical at which adverse effects have been detected while the NOAEL is the highest dose at which no adverse effects have been detected. The NOAEL (or LOAEL, if a NOAEL is not present) is adjusted downward by uncertainty factors intended to account for limitations in the available data to arrive at an exposure that is likely to be without adverse effects in humans. The NOAEL can also be compared with the human exposure estimate to derive a margin of exposure (MOE). The LOAEL and NOAEL represent an operational definition of quantities that can characterize a study, and do not necessarily have any consistent association with underlying biological processes nor with thresholds.

With the recent publication of EPA's Proposed Guidelines for Carcinogen Risk Assessment (1996a), the dichotomy between quantitative approaches for cancer and noncancer risks will begin to break down. The proposed guidelines promote the understanding of an agent's mode of action in producing tumors to reduce the uncertainty in describing the likelihood of harm and in determining the dose-response(s). The proposed guidelines call for modeling of not only

tumor data but other responses thought to be important precursor events in the carcinogenic process. The dose-response assessment under these guidelines is a two-step process: (1) response data are modeled in the range of empirical observation -- modeling in the observed range is done with biologically-based, case-specific, or with appropriate curve-fitting models; and the (2) extrapolation below the range of observation is accomplished by modeling if there are sufficient data or by a default procedure (linear, nonlinear, or both). A point of departure for extrapolation is estimated from this modeling, and low dose extrapolation proceeds from this point of departure.

The benchmark dose approach discussed in this document is a way of determining the point of departure, and can be used in cancer and noncancer risk assessment as the basis for linear low-dose extrapolation, calculation of an MOE, or application of uncertainty factors for calculating oral reference doses (RfDs), inhalation reference concentrations (RfCs) or other exposure estimates.

Several limitations to the use of the LOAEL and NOAEL have been discussed in the literature (Crump, 1984; Gaylor, 1983; Kimmel and Gaylor, 1988) and by the EPA's Science Advisory Board (EPA, 1986b, 1988a, b, 1989c). These include the fact that the NOAEL is limited to one of the doses in the study and is dependent on study design, in particular on dose selection and spacing. The NOAEL also does not account for variability in the data, and thus, a study with a limited number of animals will often result in a higher NOAEL than one which has more animals. In addition, the slope of the dose-response curve is not taken into account in the selection of a NOAEL and is not usually considered unless the slope is very steep or very shallow. Additionally, a LOAEL cannot be used to derive a NOAEL when one does not exist in a study; rather, an uncertainty factor is applied to account for this limitation.

In an effort to address some of the limitations of the LOAEL and NOAEL, Crump (1984) proposed the benchmark dose (BMD) approach as an alternative. Using this approach, the experimental data are modeled, and an oral benchmark dose (BMD) or inhalation benchmark concentration (BMC) in the observable range is estimated. The BMD/C[1] is not constrained to be

---

[1]BMD/C is used when the text can apply to values derived from either oral or inhalation exposure data.

one of the experimental doses, and can be used as a more consistent point of departure than either the LOAEL or NOAEL. This approach uses the dose-response information inherent in the data. The BMD/C accounts for the variability in the data since it is defined as the lower confidence limit on the dose estimated to produce a given level of change in response (termed the benchmark response, BMR) from controls. The BMD/C approach models all of the data in a study and the slope of the dose-response curve is integral to the BMD/C estimation. A BMD/C can be estimated even when all doses in a study are associated with a response (i.e., when there is no NOAEL). The BMD/C estimate is best when there are doses in the study near the range of the BMD/C.

B. Decision Process for Carrying out a Benchmark Dose/Concentration Analysis

This document describes the proposed approach for carrying out a complete BMD/C analysis. It is organized in the form of a decision process including the rationale and defaults for proceeding through the analysis, and follows a similar framework to that outlined in the background document (EPA, 1995c), and is meant to be used in conjunction with that document. The guidance here imposes some constraints on the BMD/C analysis through decision criteria, and provides defaults when more than one feasible approach exists. Steps in the guidance are discussed in the body of the document and in more detail in appendices to this document. The following guidance is applicable to dose-response analysis of noncancer health effects and may also be applicable to cancer dose-response analysis under the proposed guidelines (EPA, 1996a).

1. Data array analysis - endpoint selection.

Selection of the appropriate studies and endpoints is discussed in Appendix A and in various EPA publications (U.S. EPA, 1991a, 1994c, 1995f, 1996a and b). In general, the selection of endpoints to model should focus on endpoints that are relevant or assumed relevant to humans and are potentially the "critical" effect (i.e., the most sensitive). Since differences in slope could result in an endpoint having a lower BMD but a higher LOAEL or NOAEL than another endpoint, selection of endpoints should not be limited only to the one with the lowest LOAEL or NOAEL. In general, endpoints should be modeled if their LOAEL is up to 10-fold above the lowest LOAEL. This will insure that no endpoints with the potential to have the lowest

BMD are excluded from the analysis on the basis of the value of the LOAEL.

2. Minimum data set for calculating a BMD/C

Once the critical endpoints have been selected, the data sets are examined for the appropriateness of a BMD/C analysis. The following constraints on data sets are used:

- At a minimum, the number of dose groups and subjects should be sufficient to allow determination of a LOAEL.
- There must be more than one exposure group with a response different than controls (this could be determined with a pairwise comparison or a trend test). With only one responding group, there is inadequate information about the shape of the dose-response curve, and mathematical modeling becomes too arbitrary.
- Dose-response modeling is not appropriate if the responding groups all show responses near maximum (e.g., greater than 50% response for quantal data or a clear plateau of response for continuous data). In this case, there is inadequate information about the shape of the curve in the low dose region near the BMR.

3. Criteria for selecting the benchmark response (BMR)

The BMD/C approach requires that a level of change in response (the BMR) be specified in order to calculate the BMD/C. In this proposal, there are two bases for specifying the BMR: a biologically significant change in response for continuous endpoints, or the limit of detection for either quantal or continuous data. In most cases, the question concerning how much of a change in a continuous endpoint is biologically significant has not been addressed, and the level of change that is considered adverse is based in large part on the detectable level of response or limit of detection for a particular study design. For quantal data, the number of responders is counted, and again the limit of detection is used. The limit of detection is based on the background response rate, sample size, power level, and whether extra or additional risk is used in the model. Standard study designs for various endpoints should be used to determine the general background rate for a response level and the number of animals typically used. The following default decisions will be used in selection of the BMR.

- *Biological significance*: If a particular level of response for a continuous endpoint has been determined to be "biologically significant," the BMR is based on that degree of

change from background (e.g., adult body weight reduction ≥10% from the control value; a 10% or more decrease in nerve conduction velocity).

- *Limit of Detection:* When the biologically significant level of response has not been determined or is equated with a statistically significant response, the limit of detection method is used. To find the magnitude of response just detectable (the limit of detection), a default power level of 50% **(may increase, pending simulation studies)** and a one-sided test with a Type I error of 0.05 ($p<0.05$) is used.
- *Defaults:* For quantal data, a 10% increase in extra risk will be used as the default approach when neither biological significance nor the limit of detection has been determined. When the BMR is set on the basis of biological significance, extra risk will be used as a default. When the limit of detection approach is used for the BMR, it does not matter whether extra or additional risk is used, as the BMR that corresponds to the same limit of detection for the risk formulation is determined and used in the model.

4. Order of model application

The models should be executed either using software produced and distributed by the EPA or using software that carries out the curve fitting in a manner similar to the EPA software. Allowing the use of more than one model at this stage is done because there is not enough experience with modeling a wide variety of endpoints, and the best-fitting model may be somewhat endpoint-specific. The following order for running the models should be used:

- Continuous data: A linear model should be run first. If the fit of a linear model is not adequate, the polynomial model should be run, followed by the continuous power model. Other models may be applied at the discretion of the risk assessor.
- Dichotomous data: As there is currently no rationale for selecting one versus another, one or more of the following models should be applied: the log-logistic, Weibull, and quantal polynomial models. For developmental toxicity data, models with fetuses nested within litters should be used. The nested log-logistic model tended to give a better fit more often in studies by Allen et al. (1994b), but other models often gave a good fit as well. Other models may be applied at the discretion of the risk assessor.

5. Determining the model structure

The parameters included in the models most commonly used for dose-response analysis, and examination of the change in the shape of the dose-response curve as those parameters are changed, are described in the background document (EPA, 1995c). This section provides guidance on choosing a model structure appropriate to the data being analyzed:

- *Dose Parameter Restriction*: As a default, the exponent on the dose parameter in the quantal Weibull model and the coefficient on slope in the log-logistic model will be constrained to be greater than or equal to 1. Unconstrained models may be applied if necessary to fit certain data. This constraint is typically necessary to avoid unstable numerical properties in calculating the confidence interval when the parameter value is less than 1.
- *Degree of the Polynomial*: A 2-step procedure will be used as a default for determining the degree of the exponent in any polynomial model for continuous or quantal data **(this is a place-holder, pending further discussion and agreement)**. First, a default of the degree equal to k-1, where k is the number of groups, will be used. Second, a stepwise top-down reduction in the degree will be performed to select the model with the fewest parameters that still achieves adequate fit (based on $p>0.05$ from the goodness-of-fit statistic). This approach will be built into software distributed by EPA.
- *Background Parameter*: As a default approach, the background parameter will not be included in the models. A background term should be included if there is evidence of a background response. This would be the case if there is a non-zero response in the control group. A background parameter would also be included if the first 2 or more dosed groups do not show a monotonic increase in response (e.g., if the first 2 dosed groups show the same response or if the second dosed group has a lower response than the lowest dosed group). If there is doubt about whether a background parameter is needed, it is usually conservative (i.e., will result in a lower BMD) if the background parameter is excluded from the model. This is an area where more work is needed.
- *Use of extra risk versus additional risk for quantal data*: As in selection of the BMR, above.

- *Threshold Parameter*: A so-called "threshold" (intercept) term will not be included in the models used for BMD/C analysis because it is not a biologically meaningful parameter (i.e., not the same as a biological threshold) and because most data sets can be fit adequately without this parameter and the associated loss of a degree of freedom. This will be the default built into the software distributed by EPA.
- *Conversion of continuous data to dichotomous format*: The standard approach to BMD/C analysis will be to model the continuous data directly, without conversion to the dichotomous format. Alternatively, a hybrid approach such as that described by Crump (1995) can be used. The hybrid approach models the continuous data, then uses the resulting distribution of the control data to calculate a probability estimate.

  The conversion of continuous data to dichotomous data and modeling of the dichotomous data is not preferred because of the loss of information about the distribution of the response that is inherent in the approach, and because of the uncertainties in defining a cut-off response level to distinguish responders and nonresponders. Conversion to dichotomous data could be considered in the rare cases where much is known about the biological consequences of the response and a cutoff can be defined more confidently, or in cases where the need for a probabilistic estimate of response outweighs the loss of information.

6. Confidence limit calculation

The confidence limits will be calculated using likelihood theory and they will be based on the asymptotic distribution of the likelihood ratio statistic (with the exception of models fit using GEE methods, see Appendix C). This will be the default built into the software distributed by EPA. The 95% lower confidence bound on dose will be used as the default confidence interval for calculating the BMD/BMC.

7. Selection of "equivalent" models

Because each of the available models has some degree of flexibility and is capable of describing a range of dose-response patterns, it may be the case that several models seem appropriate for the analysis of the data under consideration. This section describes how to evaluate whether these models are equivalent based upon an evaluation of statistical and

biological considerations.

- All models should be retained that are not rejected at the probability of $p>0.05$ using the goodness-of-fit (GOF) statistic.
- A graphical representation of the model should also be developed and evaluated for each model. Models should be eliminated that do not adequately describe the dose-response in the range near the BMR. This is because it is possible that an adequate model fit could be obtained based on the GOF criteria alone, but the fit of the data is not adequate at the low end of the dose-response curve. Such a case should not be used for the BMD/C calculation. This is a subjective judgment and requires that software for dose-response analysis provide adequate graphical functions. No objective criteria for this decision are currently available. It should be noted, however, that most quantal models assume that each animal responds independently and has an equal probability of responding. Similarly, for continuous data, the responses are assumed to be distributed according to a normal probability distribution. When these assumptions are not appropriate, for example in studies of developmental toxicity where the responses are correlated within litters, alternative model structures may be used. Biological considerations might also be helpful to determine adequate model fit. For example, a smooth change of slope may be deemed more reasonable for a given response than an abrupt change.
- If adequate fit is not achieved because of the influence of high dose groups with high response rates, the assessor should consider adjusting the data set by eliminating the high dose group. This practice carries with it the loss of a degree of freedom, but may be useful in cases where the response plateaus or drops off at high doses. Since the focus of the BMD analysis is on the low dose and response region, eliminating high dose groups is reasonable.

At this point, the remaining models should be considered "equivalent" in terms of their usefulness for BMD/C analysis, especially when there is no biological basis for distinguishing between the models or for choosing the best model.

8. Selecting the BMD/C

As the models remaining have met the default statistical criterion for adequacy and visually

fit the data, any of them theoretically could be used for determining the BMD/C. The remaining criteria for selecting the BMD/C are somewhat arbitrary, and are adopted as defaults.

- If the BMD/C estimates from the remaining models are within a factor of 3, then they are considered to show no appreciable model dependence and will be considered indistinguishable in the context of the precision of the methods. Models are ranked based on the values of their Akaike Information Criterion (AIC), a measure of the deviance of the model fit adjusted for the degrees of freedom, and the model with the lowest AIC is used to calculate the BMD/C.
- If the BMD/C estimates from the remaining models are not within a factor of 3, some model dependence of the estimate is assumed. Since there is no clear remaining biological or statistical basis on which to choose among them, the lowest BMD/C is selected as a reasonable conservative estimate. If the lowest BMD/C from the available models appears to be an outlier, compared to the other results (e.g., if the other results are within a factor a 3), then additional analysis and discussion would be appropriate. Additional analysis might include the use of additional models, the examination of the parameter values for the models used, or an evaluation of the MLEs to determine if the same pattern exists as for the BMD/Cs. Discussion of the decision procedure should always be provided.

9. Use in Risk Assessment

The BMD/C derived based on this default approach should be used in risk assessment as discussed in other guidance documents (U.S. EPA, 1991a, 1994c, 1995f, 1996a and b).

II. INTRODUCTION

A. Background

The US EPA conducts risk assessments for an array of noncancer health effects as well as for cancer. The process of risk assessment, based on the National Research Council (NRC) (1983) paradigm, has several steps: hazard characterization, dose-response analysis, exposure assessment, and risk characterization. Risk assessment begins with a thorough evaluation of all the available data to identify and characterize potential health hazards. The next stage of the process, dose-response analysis, involves an analysis of the relationship between exposure to the chemical and health-related outcomes, and historically has been done very differently for cancer and noncancer health effects. The common practice under the 1986 Guidelines for Carcinogen Risk Assessment (EPA, 1986a) was to assume low dose linearity and model tumor incidence data by applying the linearized multistage (LMS) procedure, which extrapolates risk as the 95% upper confidence limit. This linear default position is, in part, based on the belief that the process of cancer induction is similar among chemicals, namely electrophilic reaction of carcinogens with DNA, causing mutations that are essential elements of the carcinogenic process.

The standard practice for the dose-response analysis of noncancer health effects has been to determine a lowest-observed-adverse-effect-level (LOAEL) and a no-observed-adverse-effect-level (NOAEL). The LOAEL is the lowest dose for a given chemical at which adverse effects have been detected while the NOAEL is the highest dose at which no adverse effects have been detected. The NOAEL (or LOAEL, if a NOAEL is not present) is adjusted downward by uncertainty factors intended to account for limitations in the available data to arrive at an exposure that is likely to be without adverse effects in humans. The NOAEL can also be compared with the human exposure estimate to derive a margin of exposure (MOE). The general default assumption for noncancer health effects is one of a threshold response, and thus is the basis for the NOAEL/uncertainty factor or NOAEL/MOE approach. The LOAEL and NOAEL represent an operational definition of quantities that can characterize a study, and do not necessarily have any consistent association with underlying biological processes nor with thresholds.

With the recent publication of EPA's Proposed Guidelines for Carcinogen Risk Assessment (1996a), the dichotomy between quantitative approaches for cancer and noncancer will begin to break down. The proposed guidelines promote the understanding of an agent's mode of action in producing tumors to reduce the uncertainty in describing the likelihood of harm and in determining the dose-response(s). The proposed guidelines call for modeling of not only tumor data, but other responses thought to be important precursor events in the carcinogenic process. Thus, the dose-response extrapolation procedure follows conclusions in the hazard assessment about the agent's carcinogenic mode of action. The dose-response assessment under these guidelines is a two-step process: (1) response data are modeled in the range of empirical observation -- modeling in the observed range is done with biologically-based, case-specific, or with appropriate curve-fitting models; and then (2) extrapolation below the range of observation is accomplished by modeling if there are sufficient or by a default procedure (linear, nonlinear, or both). A point of departure for extrapolation is estimated from this modeling. The linear default is a straight-line extrapolation to the origin from the point of departure, and the nonlinear default approach begins at the identified point of departure and provides a margin of exposure (MOE) analysis rather than estimating the probability of effects at low doses.

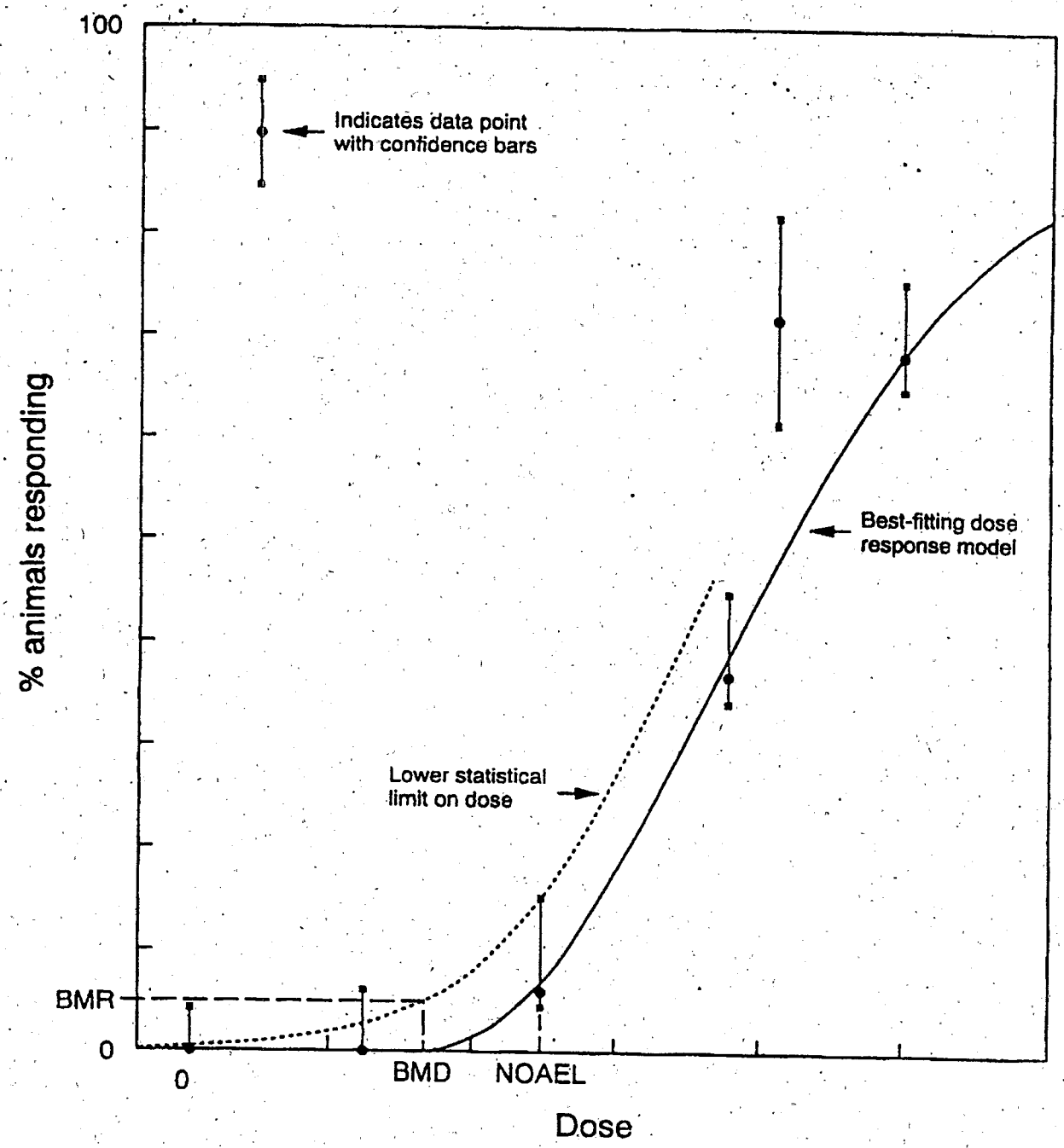
The benchmark dose approach discussed in this document is a way of determining the point of departure, and can be used as the basis for linear low-dose extrapolation, calculation of an MOE, or application of uncertainty factors for calculating oral reference doses (RfDs), inhalation reference concentrations (RfCs), or other exposure estimates.

Several limitations to the use of the LOAEL and NOAEL have been discussed in the literature (Crump, 1984; Gaylor, 1983; Kimmel and Gaylor, 1988) and by the EPA's Science Advisory Board (EPA, 1986b, 1988a,b, 1989c). These include the fact that the NOAEL is limited to one of the doses in the study and is dependent on study design, in particular on dose selection and spacing. The NOAEL also does not account for variability in the data, and thus, a study with a limited number of animals will often result in a higher NOAEL than one which has more animals. In addition, the slope of the dose-response curve is not taken into account in the selection of a NOAEL and is not usually considered unless the slope is very steep or very shallow. Additionally, a LOAEL cannot be used to derive a NOAEL when one does not exist in a study,

rather an uncertainty factor is applied to account for this limitation.

In an effort to address some of the limitations of the LOAEL and NOAEL, Crump (1984) proposed the benchmark dose (BMD) approach as an alternative. Using this approach, the experimental data are modeled, and an oral benchmark dose (BMD) or inhalation benchmark concentration (BMC) in the observable range is estimated (see Fig. 1). The BMD/C is not constrained to be one of the experimental doses, and can be used as a more consistent point of departure than either the LOAEL or NOAEL. This approach uses the dose-response information inherent in the data. The BMD/C accounts for the variability in the data since it is defined as the lower confidence limit on the dose estimated to produce a given level of change in response (termed the benchmark response, BMR) from controls. The BMD/C approach models all of the data in a study and the slope of the dose-response curve is integral to the BMD/C estimation. A BMD/C can be estimated even when all doses in a study are associated with a response (i.e., when there is no NOAEL). The BMD/C estimate is best when there are doses in the study near the range of the BMD/C.

Data used in the determination of a LOAEL/NOAEL or BMD/C are essentially of two types: quantal and continuous. Quantal data are often dichotomous (yes/no) responses and are usually presented as counts or incidence of a particular effect or of individuals affected. Quantal data are represented by such endpoints as tumor incidence, mortality, or malformed offspring. A specialized case of quantal responses is when data are categorized by severity of effect, e.g., mild, moderate, or severe. This categorization of data is often used for histopathological lesions. At the other extreme are continuous data, which represent a continuum of response and are usually represented by a measurement. Body weight, serum liver enzyme activity or nerve conduction velocity are examples of continuous responses. Continuous data can also be expressed as quantal responses by determining some magnitude of change from controls that is considered significant, then counting the number of individuals above or below that cutoff level. The level of significance can be based either on biological significance, or on statistical significance. For example, a decrease in adult body weight may not be considered adverse until it is $\geq$ 10%, this magnitude of change is based on biological significance.



BMR = target response level used to define BMD

Figure 1. Example of calculation of a BMD

Calculation of a BMD/C requires selection of a benchmark response (BMR) level, i.e., the level of change in response from considered to be an adverse effect. Basing BMRs on the level of biological significance automatically determines the level of the BMR. Basing them on statistical significance, however, is dependent on the limit of detection for a particular study design. Selection of the BMR based on the limit of detection is discussed in detail in Appendix B.

Since the initial proposal of the BMD by Crump (1984), a number of papers have been published dealing with application of the BMD/C approach. Many of these are included in the reference list for this document. A background document on the use of the BMD/C in health risk assessment was published by the EPA's Risk Assessment Forum (EPA, 1995c). This document provided a framework for the decision points in the process of applying the BMD/C approach. The background document is to be used as a companion to the current guidance document.

Several workshops and symposia have been held to discuss the application of the BMD/C and appropriate methodology (Kimmel et al., 1989; California EPA, 1993; Beck et al., 1993; SRA Symposium, 1994; Barnes et al., 1995). The most recent workshop was conducted by ILSI Risk Science Institute and sponsored by the EPA and AIHC (Barnes et al., 1995). The participants at the workshop generally endorsed the application of the BMD/C for all quantal noncancer endpoints and particularly for developmental toxicity, where a good deal of research has been done. Less information was available at the time of the EPA/AIHC/ILSI workshop on the application of the BMD/C approach to continuous data, and more work was encouraged. A number of other issues concerning the application of the BMD/C were discussed. The guidance and default options set forth in the current document are based in part on the outcome of this workshop, the background document (EPA, 1995c), and on more recent information and discussions.

A number of research efforts, many of which have dealt with reproductive and developmental toxicity data, have provided extremely useful information for application of the BMD/C approach (e.g., Alexeeff et al., 1993; Catalano et al., 1993; Chen et al., 1991; Krewski and Zhu, 1994, 1995; Auton, 1994; Crump, 1995). In a series of papers by Faustman et al. (1994), Allen et al. (1994a and b), and Kavlock et al. (1995), the BMD approach was applied to a large database of developmental toxicity studies. In brief, the results of these studies showed that

when data were expressed as counts of dichotomous endpoints (i.e., number of litters per dose group with resorptions or malformations), the NOAEL was approximately 2-3 times higher than the BMD for a 10% probability of response above control values (approximately 20 animals per dose group), and 4-6 times higher than the BMD for a 5% probability of response. When the data were expressed as the proportion of affected fetuses per litter (nested dichotomous data), the NOAEL was on average 0.7 times the BMD for a 10% probability of response, and was approximately equal, on average, to the BMD for a 5% probability of response. Expressing the data as the proportion of affected fetuses per litter is the more appropriate way to analyze developmental toxicity data. However, the results of the quantal data analysis also may apply to using the BMD/C approach with other quantal data, and suggest that the NOAEL in these cases may be at or above the 10% true response level, depending on sample size and background rate.

Since reduced fetal weight in developmental toxicity studies often shows the lowest NOAEL among the various endpoints evaluated, the application of the BMD to these continuous data also was evaluated (Kavlock et al., 1995). A variety of cutoff values was explored for defining an adverse level of weight reduction below control values. In some cases, data were analyzed using a continuous power model, and in other cases, the data were transformed to dichotomous data. Comparisons with the NOAEL showed that several cutoff values could be used to give values similar to the NOAEL. These analyses suggest ways in which BMD/Cs may be developed for continuous data from a variety of endpoints.

In a recent paper, Crump (1995) detailed a new approach to deriving a BMD/C for continuous data based on a method originally proposed by Gaylor and Slikker (1990). This approach makes use of the distribution of continuous data, estimates the incidence of individuals falling above or below a level considered to be adverse, and gives the probability of responses at specified doses above the control levels. This results in an expression of the data in the same terms as that derived from analyses of quantal data, and allows more direct comparison of BMDs derived from continuous and quantal data. This approach has not been applied to many data sets as software has not been developed until recently.

Another approach to the derivation of BMD/Cs for noncancer health effects is the multinomial modeling approach. Many noncancer health effects are characterized by multiple

endpoints that are not completely independent of one another. Catalano et al. (1993), Chen et al. (1991), and Zhu et al. (1994) have worked on this issue using developmental toxicity data, and have shown that, in general, the BMD derived from a multinomial modeling approach is lower than that for any individual endpoint. This approach has not been applied to other health effects data, but should be kept in mind when multiple related outcomes are being considered for a particular health effect.

Other types of approaches have been proposed for analysis of noncancer data, e.g., categorical regression modeling (Dourson et al., 1985; Hertzberg, 1989; Hertzberg and Miller, 1985; Guth; Simpson), and Bayesian approaches (Hasselblad and Jarabek, 1995), that may be more appropriate in some cases. There has not been much experience as yet in using these approaches for deriving BMD/Cs.

As discussed earlier most of the effort on development of BMD/C approaches has been related to noncancer health effects, but recent proposed guidelines for cancer risk assessment (EPA, 1996a) suggest an approach to dose-response analysis that is similar to use of the BMD/C for noncancer risk assessment. While terminology and exact procedures are not the same, the concept is parallel to the BMD/C approach (see Section V for further discussion). Thus, guidance is provided in this document on the use of the BMD/C as the point of departure for low dose extrapolation of both cancer and noncancer health effects.

B. Purpose of This Guidance Document

The purpose of this document is to provide guidance for the Agency and the outside community on the application of the BMD/C approach in the calculation of RfD/Cs and other exposure standards and in the estimation of cancer risks for EPA use. The document provides guidance based on today's knowledge and understanding, and on experience gained in using this approach. The Agency is actively applying this methodology and evaluating the outcomes for the purpose of gaining experience in using it with a variety of endpoints. This document is intended to be updated as new information becomes available that would suggest approaches and default options alternative or additional to those indicated here. The document should not be viewed as precluding additional research on modified or alternative approaches that will improve

quantitative risk assessment. In fact, the use of improved scientific understanding and development of more mechanistically-based approaches to dose-response modeling is strongly encouraged by the Agency.

C. Definition of the Benchmark Dose or Concentration

The benchmark dose or concentration (BMD/C) is defined as the statistical lower confidence limit on the dose estimated to produce a predetermined level of change in response (the benchmark response - BMR) relative to controls.

The BMD/C is intended to be used as an alternative to the NOAEL in deriving a point of departure for low dose extrapolations. The BMD/C is a dose corresponding to some change in the level of response relative to background, and is not dependent on the doses used in the study. The BMR is based on a biologically significant level of response or on the response level at the lower detection limit of the observable dose range for a particular endpoint in a standard study design. The BMD/C approach does not reduce uncertainty inherent in extrapolating from animal data to humans (except for that in the LOAEL to NOAEL extrapolation), and does not require that a study identify a NOAEL, only that at least one dose be near the range of the response level for the BMD/C.

III. BENCHMARK DOSE GUIDANCE

This section describes the proposed approach for carrying out a complete BMD/C analysis. It is organized in the form of a decision process including the rationale and defaults for proceeding through the analysis, and follows a similar framework to that outlined in the background document (EPA, 1995c). The guidance here imposes some constraints on the BMD/C analysis through decision criteria, and provides defaults when more than one feasible approach exists. Steps in the guidance are discussed here and in more detail in Appendices A-C.

A. Data Array Analysis - Endpoint Selection

The first step in the process is a complete qualitative review of the literature to identify and characterize the hazards related to a particular compound or exposure situation. This process is the same whether a BMD/C analysis or a NOAEL approach is used. Guidance on review of data for risk assessment can be found in a number of EPA publications (EPA, 1991a, 1994c, 1995f, 1996a and b). Further discussion of design issues, data reporting, and route extrapolation relevant to the BMD/C analysis can be found in Appendix A.

Following a complete qualitative review of the data, the risk assessor must select the studies appropriate for benchmark dose analysis. The selection of the appropriate studies is based on the human exposure situation that is being addressed, the quality of the studies, and the relevance and reporting adequacy of the endpoints.

The process of selecting studies for benchmark analysis is intended to identify those studies for which modeling is feasible, so that BMD/Cs can be calculated and used in risk assessment. In most cases the selection process will identify a single study or very few studies for which calculations are relevant, and all studies should be modeled. However, cases with many studies, or studies in which many endpoints are reported would require a very large number of BMD/C calculations. Multivariate analysis of such studies to describe correlations between effects is a useful tool, but has not been explored in terms of the BMD/C method.

The analysis of a large number of endpoints may result in benchmark calculations which are redundant in the information they convey. In these cases, it is useful to select a subset of the

endpoints as representative of the effects in the target organ or the study. This selection can be made on the basis of sensitivity or severity, which may be more easily compared within a single study in the same target organ than across studies. Within an experiment, an endpoint may be selected based on how well it represents others for the same target organ and on its dose-response behavior. It is reasonable to select a representative endpoint that shows smoothly increasing response with increasing dose in order to obtain a good fit of the dose response model.

1. Selection of Endpoints to be Modeled.

Once endpoints have been evaluated with regard to their relevance for BMD/C analysis, the selection of endpoints to model should focus on endpoints that are relevant or assumed relevant to humans and potentially the "critical" effect (i.e., the most sensitive). Since differences in slope could result in an endpoint having a lower BMD but a higher LOAEL or NOAEL than another endpoint, selection of endpoints should not be limited only to the one with the lowest LOAEL. In general, endpoints should be modeled if their LOAEL is up to 10-fold above the lowest LOAEL. This will insure that no endpoints with the potential to have the lowest BMD/C are excluded from the analysis on the basis of the value of the LOAEL.

2. Minimum Data Set for Calculating a BMD/C

Once the critical endpoints have been selected, the data sets are examined for the appropriateness of a BMD/C analysis. The following constraints on data sets are used:

- At a minimum, the number of dose groups and subjects should be sufficient to allow determination of a LOAEL.
- There must be more than one exposure group with a response different than controls (this could be determined with a pairwise comparison or a trend test). With only one responding group, there is inadequate information about the shape of the dose-response curve, and mathematical modeling becomes too arbitrary.
- Dose-response modeling is not appropriate if the responding groups all show responses near maximum (e.g., greater than 50% response for quantal data or a clear plateau of response for continuous data). In this case, there is inadequate information about the shape of the curve in the low dose region near the BMR.

3. Combining Data for a BMD/C Calculation

Data sets that are statistically and biologically compatible may be combined prior to dose-response modeling and thus generate fewer BMD/C estimates for comparison. The combining of appropriate data sets prior to modeling has several advantages. It leads to increased confidence, both statistical and biological, in the calculated BMD/C. In addition, the use of combined data sets may encourage further research to be conducted on that compound if the additional data can affect the BMD/C estimate. Example #3 for boron in Appendix D is a case where data could be combined for the BMD/C analysis.

B. Criteria for Selecting the Benchmark Response Level (BMR)

The Benchmark Dose Workshop (Barnes et al, 1995) recommended that the BMR "should be within or near the experimental range of the doses studied." The Workshop's recommendation in this regard was to use a 5% or 10% increase in response levels above controls as the BMR. However, this recommendation addressed only the case of dichotomous data and did not deal with continuous endpoints. In addition, the variety of experimental designs used in toxicology and the associated variety of experimental endpoints and their ranges of variability may require a broader range of approaches. This proposal for selecting the BMR attempts to take into account the wide array of toxicological responses.

Any toxicological study is limited in its ability to detect an increase in the incidence of adverse responses or a change in the mean of continuous endpoints. This limitation is determined by aspects of the study design: largely, the number of independent experimental units per dose group, the nature of any nesting or repeated measures in the design, the background incidence of the adverse response (for dichotomous endpoints) or the variance of the control response (for continuous endpoints), as well as the intended form of significance testing. For some endpoints, such as adult weight, serum liver enzyme activities, and certain neurological measurements like nerve conduction velocity, there is also a minimum change from control levels that is considered "biologically significant." For endpoints for which there is no agreed upon biologically significant change, quantifying the concept of "limit of detection" for a toxicological study can simplify specifying the BMR.

The BMD/C approach requires that a level of change in response (the BMR) be specified in order to calculate the BMD/C. In this proposal, there are two bases for specifying the BMR: a biologically significant change in response for continuous endpoints, or the limit of detection for either quantal or continuous data. In most cases, the question concerning how much of a change in a continuous endpoint is biologically significant has not been addressed, and the level of change that is considered adverse is based in large part on the detectable level of response or limit of detection for a particular study design. For quantal data, the number of responders is counted, and again the limit of detection is used. The limit of detection is based on the background response rate, sample size, power level, and whether extra or additional risk is used in the model. Standard study designs[2] for various endpoints should be used to determine the general background rate for a response level and the number of animals typically used. These concepts are discussed in depth in Appendix B.

The following default decisions will be used in selection of the BMR.

- *Biological significance*: If a particular level of response for a continuous endpoint has been determined to be "biologically significant," the BMR is based on that degree of change from background (e.g., adult body weight reduction ≥10% from the control value; a 10% or more decrease in nerve conduction velocity).
- *Limit of Detection:* When the biologically significant level of response has not been determined or is equated with a statistically significant response, the limit of detection method is used. To find the magnitude of response just detectable (the limit of detection), a default power level of 50% **(may increase, pending simulation studies)** and a one-sided test with a Type I error of 0.05 ($p<0.05$) is used.
- *Defaults:* For quantal data, a 10% increase in extra risk will be used as the default approach when neither biological significance nor the limit of detection has been

[2]Using a cadre of standard study designs for a variety of endpoints is one way to reduce the uncertainty that can be introduced when different individuals make decisions for limits of detection. We assume here the use of Agency testing protocols as the basis for standard "good" study designs. In areas where standard testing protocols have not been developed, the Agency encourages activities that can assist in identifying a "most common well-designed protocol" for typically-studied endpoints.

determined. When the BMR is set on the basis of biological significance, extra risk will be used as a default. When the limit of detection approach is used for the BMR, it does not matter whether extra or additional risk is used, as the BMR that corresponds to the same limit of detection for the risk formulation is determined and used in the model.

C. Mathematical Modeling

1. Introduction

The goal of the mathematical modeling in benchmark dose computation is to fit a model to dose-response data that describes the data set, especially at the lower end of the observable dose-response range. The fitting must be done in a way that allows the uncertainty inherent in the data to be quantified and related to the estimate of the dose that would yield the benchmark response. In practice, this procedure will involve first selecting a family or families of models for further consideration, based on characteristics of the data and experimental design, and fitting the models using one of a few established methods. Subsequently, a lower bound on dose is calculated at the BMR. This section is too brief to do more than introduce the topic of nonlinear modeling. Some references for further reading are: Chapter 10 of Draper and Smith (1981), Clayton and Hills (1993), Davidian and Giltinan (1995), and McCullagh and Nelder (1989).

Dose-response models are expressed as functions of dose, possibly covariates, and a set of constants, called parameters, that govern the details of the shape of the resulting curve. They are fitted to a data set by finding values of the parameters that adjust the predictions of the model for observed values of dose and covariates to be close to the observed response. At the present, although biological models may often be expressed as nonlinear models (e.g., Michaelis-Menton processes), nonlinear models do not necessarily have a biological interpretation. Thus, criteria for final model selection will be based solely on whether various models describe the data, conventions for the particular endpoint under consideration, and, sometimes, the desire to fit the same basic model form to multiple data sets. Since it is preferable to use special purpose modeling software, EPA is in the process of developing user-friendly software which includes several models and default processes as described in this document. The models included in the software are listed in Appendix E.

This section provides guidance on how to go about choosing a model structure appropriate to the data being analyzed, the order of model application confidence limit calculation, selection of "equivalent" models, and selection of the BMD/C to use as the point of departure. More in depth discussion of these topics can be found in Appendix C. Examples of BMD/C modeling of actual data sets can be found in Appendix D.

2. Order of Model Application

The models should be executed either using software produced and distributed by the EPA or using software that carries out the curve fitting in a manner similar to the EPA software. Allowing the use of more than one model at this stage is done because there is not enough experience with modeling a wide variety of endpoints, and the best-fitting model may be somewhat endpoint-specific. The following order for running the models should be used:

- Continuous data: A linear model should be run first. If the fit to a linear model is not adequate, the polynomial model should be run, followed by the continuous power model. Other models may be applied at the discretion of the risk assessor.
- Dichotomous data: As there is currently no rationale for selecting one versus another, one or more of the following models should be applied: the log-logistic, Weibull, and polynomial models. For developmental toxicity data, models with fetuses nested within litters should be used. The nested log-logistic model tended to give a better fit more often in studies modeled by Allen et al. (1994b), but other models often gave a good fit as well. Other models may be applied to dichotomous data at the discretion of the risk assessor.

3. Determining the Model Structure

The parameters included in the models most commonly used for dose-response analysis, and examination of the change in the shape of the dose-response curve as those parameters are changed, are described in the background document (EPA, 1995c). This section provides guidance on choosing a model structure appropriate to the data being analyzed.

- *Dose Parameter Restriction*: As a default, the exponent on the dose parameter in the quantal Weibull model and the coefficient on slope in the log-logistic model will be constrained to be greater than or equal to 1. Unconstrained models may be applied if necessary to fit certain data. This constraint is typically necessary to avoid unstable

numerical properties in calculating the confidence interval when the parameter value is less than 1.

- *Degree of the Polynomial*: A 2-step procedure will be used as a default for determining the degree of the exponent in the polynomial model for both continuous and quantal data **(this is a place-holder, pending further discussion and agreement)**. First, a default of the degree equal to k-1, where k is the number of groups, will be used. Second, a stepwise top-down reduction in the degree will be performed to select the model with the fewest parameters that still achieves adequate fit (based on p>0.05 from the goodness-of-fit statistic). This approach will be built into software distributed by EPA.
- *Background Parameter*: As a default approach, the background parameter will not be included in the models. A background term should be included if there is evidence of a background response. This would be the case if there is a non-zero response in the control group. A background parameter would also be included if the first 2 or more dosed groups do not show a monotonic increase in response (e.g., if the first 2 dosed groups show the same response or if the second dosed group has a lower response than the lowest dosed group). If there is doubt about whether a background parameter is needed, it is usually conservative (i.e., will result in a lower BMD) if the background parameter is excluded from the model. This is an area where more work is needed.
- *Use of extra risk versus additional risk for quantal data*: As in selection of the BMR, above.
- *Threshold parameter*: A so-called "threshold" (intercept) term will not be included in the models used for BMD/C analysis because it is not a biologically meaningful parameter (i.e., not the same as a biological threshold) and because most data sets can be fit adequately without this parameter and the associated loss of a degree of freedom. This will be the default built into the software distributed by EPA.
- *Conversion of continuous data to dichotomous format*: The standard approach to BMD/C analysis will be to model the continuous data directly, without conversion to the dichotomous format. Alternatively, a hybrid approach such as that described by Crump (1995) can be used. The hybrid approach models the continuous data, then uses the

resulting distribution of the control data to calculate a probability estimate.

The conversion of continuous data to dichotomous data and modeling of the dichotomous data is not preferred because of the loss of information about the distribution of the response that is inherent in the approach, and because of the uncertainties in defining a cut-off response level to distinguish responders and nonresponders. Conversion to dichotomous data could be considered in the rare cases where much is known about the biological consequences of the response and a cutoff can be defined more confidently, or in cases where the need for a probabilistic estimate of response outweighs the loss of information.

4. Confidence Limit Calculation

The confidence limits will be calculated based using likelihood theory and they will be based on the asymptotic distribution of the likelihood ratio statistic (with the exception of models fit using GEE methods; see Appendix C). This will be the default built into the software distributed by EPA. The 95% lower confidence bound on dose will be used as the default confidence interval for calculating the BMD/BMC.

5. Selection of "Equivalent" Models

Because each of the available models has some degree of flexibility and is capable of describing a range of dose-response patterns, it may be the case that several models seem appropriate for the analysis of the data under consideration. This section describes how to evaluate whether these models are equivalent based upon an evaluation of statistical and biological considerations.

- All models should be retained that are not rejected at the probability of $p>0.05$ using the goodness-of-fit (GOF) statistic.
- A graphical representation of the model should also be developed and evaluated for each model. Models should be eliminated that do not adequately describe the dose-response in the range near the BMR. This is because it is possible that an adequate model fit could be obtained based on the GOF criteria alone, but the fit of the data is not adequate at the low end of the dose-response curve. Such a case should not be used for the BMD/C calculation. This is a subjective judgment and requires that software for dose-response

analysis provide adequate graphical functions. No objective criteria for this decision are currently available. It should be noted, however, that most quantal models assume that each animal responds independently and has an equal probability of responding. Similarly, for continuous data, the responses are assumed to be distributed according to a normal probability distribution. When these assumptions are not appropriate, for example in studies of developmental toxicity where the responses are correlated within litters, alternative model structures may be used. Biological considerations might also be helpful to determine adequate model fit. For example, a smooth change of slope may be deemed more reasonable for a given response than an abrupt change.

- If adequate fit is not achieved because of the influence of high dose groups with high response rates, the assessor should consider adjusting the data set by eliminating the high dose group. This practice carries with it the loss of a degree of freedom, but may be useful in cases where the response plateaus or drops off at high doses. Since the focus of the BMD analysis is on the low dose and response region, eliminating high dose groups is reasonable.

At this point, the remaining models should be considered "equivalent" in terms of their usefulness for BMD/C analysis, especially when there is no biological basis for distinguishing between the models or for choosing the best model.

6. Selecting the BMD/C

As the models remaining have met the default statistical criterion for adequacy and visually fit the data, any of them theoretically could be used for determining the BMD/C. The remaining criteria for selecting the BMD/C are necessarily somewhat arbitrary, and are adopted as defaults.

- If the BMD/C estimates from the remaining models are within a factor of 3, then they are considered to show no appreciable model dependence and will be considered indistinguishable in the context of the precision of the methods. Models are ranked based on the values of their Akaike Information Criterion (AIC), a measure of the deviance of the model fit adjusted for the degrees of freedom, and the model with the lowest AIC is used to calculate the BMD/C.
- If the BMD/C estimates from the remaining models are not within a factor of 3, some

model dependence of the estimate is assumed. Since there is no clear remaining biological or statistical basis on which to choose among them, the lowest BMD/C is selected as a reasonable conservative estimate. If the lowest BMD/C from the available models appears to be an outlier, compared to the other results (e.g., if the other results are within a factor a 3), then additional analysis and discussion would be appropriate. Additional analysis might include the use of additional models, the examination of the parameter values for the models used, or an evaluation of the MLEs to determine if the same pattern exists as for the BMD/Cs. Discussion of the decision procedure should always be provided.

- In some cases, relevant data for a given agent are not amenable to modeling and a mixture of BMD/Cs and NOAEL/LOAELs results. When this occurs, and the critical effect is from a study considered adequate but not amenable to modeling, the NOAEL should be used as the point of departure.

IV. USING THE BMD/C IN NONCANCER DOSE-RESPONSE ANALYSIS

A. Introduction

Use of the BMD/C approach for noncancer dose-response analysis on a broad scale raises a number of issues. A wide variety of endpoints of noncancer health effects must be considered in the development of exposure limits. Experience with the BMD/C approach varies depending on the disciplinary area and type of endpoint under consideration. As indicated earlier in the document, a good deal of work has been done for developmental toxicity, less for neurotoxicity, and relatively little for other types of noncancer health effects to date. Although some BMD/Cs are based on continuous variables and others on discrete (mostly dichotomous) variables, this is not a problem using the approach advocated in this document. Basically, the BMR is set by selecting levels of response that are considered biologically significant and/or near the limit of detection for a particular endpoint in a given study design, using much the same basis as has been done for setting NOAELs.

A few RfCs and one RfD included in the IRIS database (EPA, 1996c) have been based on BMD/C calculations. These include: methyl mercury (RfD based on delayed postnatal development in humans), carbon disulfide (Example #1 in Appendix D), HFC-134a (Example # 2 in Appendix D) and antimony trioxide (RfC based on chronic pulmonary interstitial inflammation in female rats). A few other risk assessments have been developed using the BMD/C approach, including manganese (RfC based on neurotoxicity in humans) (EPA, 1995g), and diesel exhaust (RfC based on lung irritation) (EPA, 1995), and others are under development (e.g., boron - Example #3 in Appendix D; 1,3-butadiene - Example #4 in Appendix D). All of these cases have involved a variety of approaches to calculating the BMD/C and to deriving RfDs and/or RfCs. In addition, the Agency is continuing to conduct such evaluations as part of ongoing risk assessments and is aware of a number of similar efforts outside the Agency.

B. Effect of the BMD/C Approach on Use of Uncertainty Factors

Uncertainty always accompanies inferences regarding the derivation of RfD/Cs or other exposure limits. Whether a threshold for noncancer effects is assumed or not, inferences about

lower human exposure levels usually embody one or more extrapolations which may include: animal to human, human to sensitive human populations, acute exposure to chronic exposure, and LOAEL to NOAEL (if a NOAEL is not available). Each of these extrapolations carries attendant uncertainties. The only change in the use of uncertainty factors when using the BMD/C as the point of departure is that the LOAEL to NOAEL uncertainty factor is no longer necessary.

C. Dose-Response Characterization

Dose-response analyses based on BMD/Cs provide a point of departure for calculating MOEs, RfD/Cs, and other exposure estimates. Although the BMD/C is associated with a defined level of risk in the study population from which the BMD/C was calculated, it would be misleading to translate that to a level of risk at the MOE or RfD/C or other reference value. The dose at the MOE or reference value is intended to yield human risks that are lower than the risk to test species at the BMD/C, but the exact degree of protection is unknown, given the attendant uncertainties in extrapolation of low dose nonlinear response relationships.

This issue was discussed in some detail at the BMD workshop (Barnes et al., 1995). The participants concluded that the level of effect at the BMD/C in the experimental species cannot be translated to a level of risk at the RfD/C for humans. They pointed out that typically the average level of effect in the test species at the BMD/C is less than the level of response at the BMR (e.g., 5%, 10%), since the BMD/C is defined as the lower 95% confidence limit on the dose at the BMR. They recommended an approach (discussed further below) to communicating the use of the BMD/C in such situations, and that this information be included in IRIS along with the central estimates (e.g., MLEs), BMD/Cs, MOEs, RfD/Cs or other reference values.

The following statement, modified from the BMD workshop report, is an example of what can be used to communicate these values based on BMD/Cs:

> The BMD/C corresponds to a dose level which yields (with 95% confidence) a level of effect in the test species of, for example, 10% or less for quantal data, or that represents a change from the control mean of, for example, 5% for continuous data. This is about the lowest level

of effect that can be detected reliably in an experimental study of this design. Alternatively, if the BMR is based on a degree of change that is considered biologically significant, e.g., a 10% or greater reduction in adult body weight or reduction in birth weight below 2.5 kg, the BMD/C represents the lower confidence limit on dose for that degree of change. Overall, the BMD/C will be a more consistent point of departure than the NOAEL and will not be constrained by the doses used in a particular study.

V. USE OF BENCHMARK-STYLE APPROACHES IN CANCER RISK ASSESSMENT

Although use of a benchmark approach in noncancer assessment is relatively new, benchmark-style approaches have been used in ranking and regulating potential carcinogens for many years. Historically, cancer risk assessors have not used the term "benchmark dose," preferring terms such as "TD_{50}" and "ED_{10}," because these latter terms draw an analogy to the LD_{50} and maintain continuity in terminology with the published cancer literature. The principal use these approaches has been in hazard ranking; more recently, a benchmark-style approach is being proposed for extrapolation below the observed range for tumor incidence in the proposed cancer guidelines (EPA, 1996a). This section provides guidance for cancer hazard ranking and low-dose extrapolation.

A. Use in Hazard Ranking

Hazard ranking, the comparison of quantitative potency estimates across chemicals, has been the principal use of ED_{10}s and similar estimates. This use does not require that low-dose risk be quantified, consequently, rankings can be based on potency estimates from the experimental range, which are mostly independent of choice of model. EPA has used ED_{10}s in regulations and proposals that set priorities for emergency response (U.S. EPA, 1987, 1989a, 1995a) and evaluate health hazards from air pollutants (U.S. EPA, 1994a). This section provides guidance for developing ED_{10}s, cites sources where TD_{50}s and ED_{10}s can be obtained, and reviews examples where TD_{50}s and ED_{10}s have been adapted for other uses.

1. Guidance for Developing Comparable ED_{10}s

Hazard rankings are based on the notion that potency estimates for different chemicals and experimental protocols can be made comparable. It is customary to adjust for background tumor rates, combine fatal and incidental tumors, and correct for early mortality (Peto et al., 1984; Sawyer et al., 1984). Other calculation conventions include a 2-year standard lifespan for rat and mouse studies; standard food, water, and air intake factors for each sex and species; standard absorption rates for each exposure route; use of time-weighted-average doses; and correction for less-than-lifespan studies (Peto et al., 1984). A common method promotes consistency; for

example, lifetable methods give similar, but usually lower, TD_{50}s than summary incidence methods (Gold et al., 1986a).

These factors are covered in EPA's methodology for developing ED_{10}s and applying them in hazard rankings (EPA, 1988c, 1994b), incorporated here by reference. The methodology specifies detailed approaches for selecting data sets and tumor responses; deriving equivalent doses across species, dosing regimens, and exposure routes; adjusting for less-than-lifetime dosing or survival; and modeling in the experimental range. To allow consistent estimates where data are missing, the methodology provides defaults. With this approach, ED10s for different chemicals and experimental protocols are made comparable.

2. Sources of TD_{50}s and ED_{10}s

TD_{50}s and ED_{10}s are readily available from the published literature. The first and most comprehensive compilation is the Carcinogenic Potency Database (CPDB), developed and regularly updated by Gold et al. (1984, 1986b, 1987, 1990, 1993, in press). The CPDB provides standardized information on over 4400 carcinogenicity experiments on over 1100 chemicals. Carcinogenic potency is described by the TD_{50}, defined as "the chronic dose rate that will halve the probability of remaining tumor-free throughout the standard life span." TD50s have been used to investigate many questions, including those concerning chemical carcinogenesis, carcinogen identification, cross-species extrapolation, reproducibility of results, and ranking possible carcinogenic hazards. TD_{50}s from the CPDB have been used by California EPA as an alternative method for low-dose extrapolation (Hoover et al., 1995).

EPA uses ED_{10}s to rank potential carcinogens found at waste sites (EPA, 1988c, 1989b, 1995b) and evaluate health hazards from air pollutants (EPA, 1994b). These references provide standardized cancer potency estimates for more than 100 chemicals found at waste sites and more than 80 air pollutants. ED_{10}s from these references have been used in other comparative risk analyses. The 10-percent level was chosen to represent the lower end of the experimental range and, thus, be more pertinent to human environmental exposure than TD_{50}s.

Judgment is essential when adapting TD_{50}s or ED_{10}s to a particular application. For example, the inclusive nature of the CPDB supports use of TD_{50}s as a midrange response, as many carcinogenicity experiments test only one high dose. In contrast, EPA uses ED_{10}s from the lower

end of the experimental range because of its focus on human environmental exposure.

B. Use in Low-Dose Extrapolation

Low-dose extrapolation goes beyond the experimental information on tumor incidence to estimate risks at lower doses. Extrapolations generally have been described by a slope factor, which is an upper bound on the slope of an assumed linear dose-response curve at low doses. Slope factors can be multiplied by exposure levels to bound the cancer risk. Recently proposed cancer guidelines (EPA, 1996a) provide alternative approaches to low-dose extrapolation, with linear, nonlinear or both as defaults based on an LED_{10}. This section summarizes the proposed use of LED_{10}s, describes how the proposed use of LED_{10}s would change existing slope factors, and discusses some issues that have appeared in the literature.

1. Proposal for low-dose extrapolation under the 1996 Proposed Guidelines for Carcinogen Risk Assessment

As described previously, the dose-response assessment under the new guidelines is a two-step process. In the first step, response data are modeled in the range of observation. It should be noted that in addition to modeling tumor data, the proposed guidelines allow for the opportunity to use and model other kind of responses if they are considered to be important measures f carcinogenic risk. In the second step, extrapolation below the range of observation is accomplished by biologically based or case-specific modeling if there are sufficient data, or by a default procedure using a curve-fitting model. The proposed guidelines indicate a preference for biologically based dose-response models, such as the two-stage model of initiation plus clonal expansion and progression (Moolgavkar and Knudson, 1981; Chen and Farland, 1991; EPA, 1995d) for the extrapolation of risk. Because the parameters of these models require extensive data, it is anticipated that the necessary data to support these models will not be available for most chemicals. Therefore, the 1996 proposed guidelines allow the use of several default extrapolation approaches.

The default extrapolation approaches are based on "curve-fitting" in the observed range to

determine the lower 95% confidence limit on a dose associated with 10% extra risk (LED_{10}[3]). The LED_{10} is proposed as a standard point of departure. The 10% response is at or near the limit of sensitivity in most cancer bioassays. Other points of departure may be appropriate, e.g., if a response is observed below the 10% level. The point of departure forms the basis for the default extrapolation approaches described below.

Linear default extrapolation procedure--The LMS procedure of the 1986 guidelines for extrapolating risk from upper confidence intervals is no longer recommended as the linear default in the 1996 proposed guidelines. The linear default in the new guidelines is a straight-line extrapolation to the origin from the point of departure identified by curve fitting in the range of observed data (the slope of this line is $0.10/LED_{10}$, which is inversely related to the LED_{10}, i.e., high potency is indicated by a large slope and a small LED_{10}). It should be noted that the straight-line extrapolation from the LED_{10} and the LMS procedure produce similar results (Gaylor and Kodell, 1980).

The straight-line/LED_{10} approach does not imply unfounded sophistication as extrapolation with the LMS procedure does. The linear default approach would be considered for agents that directly affect growth control at the DNA level (e.g., carcinogens that directly interact with DNA and produce mutations). There might be modes of action other than DNA reactivity (e.g., certain receptor-based mechanisms) that are better supported by the assumption of linearity. When inadequate or no information exists to explain the mode of action of a carcinogen, the linear default approach would be used as a science policy choice in the interest of public health. Likewise, a linear default would be used if evidence demonstrates the lack of support for linearity (e.g., negative genotoxicity studies), but there is also an absence of sufficient information on another mode of action to explain the induced tumor response. The latter is also a public health conservative policy choice.

Nonlinear default extrapolation procedure--Although the understanding of the mechanisms of induced carcinogenesis likely will never be complete for most agents, there are situations where

---

[3]95 percent lower bound on the effective dose associated with a 10 percent increase in response; equivalent to the BMD/C. The term LED_{10} is used in this section becuase of its use in the proposed guidelines for carcinogen risk assessment (EPA, 1996a).

evidence is sufficient to support an assumption of nonlinearity. Because it is experimentally difficult to distinguish modes of actions with true "thresholds" from others with a nonlinear dose-response relationship, the proposed nonlinear default procedure is considered a practical approach to use without the necessity of distinguishing sources of nonlinearity. Moreover, the use of empirical models to approximate a nonlinear dose-response relationship is discouraged, because different models can give dramatically different results below the observable range of tumor response, with no basis for choosing among them. Thus, in the 1996 proposed guidelines, the nonlinear default approach begins at the identified point of departure (LED_{10}) and provides a margin of exposure (MOE) analysis rather than estimating the probability of effects at low doses.

A nonlinear default position must be consistent with the understanding of the agent's mode of action in causing tumors. For example, a nonlinear default approach would be taken for agents causing tumors as a secondary consequence of organ toxicity or induced physiological disturbances. Because there must be a sufficient understanding of the agent's mode of action to take the nonlinear default position, and because the proposed guidelines allow for the opportunity to model not only tumor data but other responses thought to be important precursor events in the carcinogenic process (e.g., DNA adducts, mutation, cellular proliferation, hormonal or physiological disturbances, receptor binding), modeling of key nontumor data is anticipated to make extrapolation based on the nonlinear default procedures more meaningful by providing insights into the relationships of exposure and tumor response below the observable range. Nontumor data may actually be used instead of tumor data for determining the point of departure for the MOE analysis.

The MOE analysis is used to compare the LED_{10} with the human exposure levels of interest. The acceptability of an MOE is a matter for risk management; thus, the key objective of the MOE analysis is to describe for the risk manager how rapidly response may decline with dose. The MOE analysis considers:

- **steepness of the slope** of the dose response, including the degree to which the dose-response relationship deviates from a straight line.
- **human differences in sensitivity** - if this cannot be determined it is considered to be at least 10-fold.

- **interspecies differences** - if this cannot be determined, humans can be considered 10-fold more sensitive; if evidence shows humans to be less sensitive, a fraction no smaller than 1/10 can be used. This compares cross-species sensitivity to equivalent doses, which are calculated using either toxicokinetic information or an oral default scaling factor based on equivalence of $mg/kg^{3/4}$-d (US EPA, 1992), or inhalation default using the RfC dosimetry approach (USEPA, 1994c).
- **nature of the response** being used for the point of departure, i.e., tumor or nontumor data - tumor data might support a greater MOE than a more sensitive precursor response which can be measured at lower exposures.
- **biopersistence of the agent** becomes an important factor in the MOE analysis if nontumor precursor response data from less than life-time exposures are used for determining the point of departure for extrapolation of risk.

Both linear and nonlinear defaults--There may be situations where it is appropriate to consider both linear and nonlinear default procedures. For example, an agent may produce tumors at multiple sites by different mechanisms. If it is apparent that an agent is both DNA reactive and highly active as a promoter at higher doses, both linear and nonlinear default procedures may be used to distinguish between the events operative at different portions of the dose-response curve and to consider the contribution of both phenomena.

In curve-fitting the data in the observed range to determine a point of departure for the defaults discussed above, the multistage model can be used as a default empirical curve-fitting model; personal computer software has been developed for EPA analyses (Howe et al., 1986). Results from other software packages, even those based on the same model, can be difficult to compare, because they can employ different parameter constraints, dose transformations, background incidence adjustments, confidence interval methods, or goodness of fit statistics. For example, a spreadsheet version of the multistage model (Haas, 1994) uses a logarithmic transformation of dose to implement the nonnegativity constraints in the multistage model; such variants can be statistically sound yet introduce small differences in results compared to other EPA analyses.

Slope factors developed from LED_{10}s and alternative points of departure, including ED_{10}s

and ED_{01}s, are stable in the face of minor perturbations in incidence in controls and at the lowest tumor dose (Gaylor et al., 1994). This represents an improvement over the maximum likelihood estimate of slope from the linearized multistage procedure, which is not statistically stable below the experimental range (Krewski et al., 1990, 1991; Gaylor et al., 1994).

[Note: Use of nontumor data is currently being discussed, and additional information will be added later.]

2. Issues in choosing a point of departure

The approach of the proposed cancer guidelines (U.S. EPA, 1996a) reflects an observation that carcinogenicity bioassay studies can provide information in the experimental range, but cannot determine the nature of the true dose-response relationship below that range; this suggests the separation of analysis in the experimental range from extrapolation to lower doses (Gaylor and Kodell, 1980). Using a model to extrapolate to a pre-determined risk level, with further extrapolation by straight line, was first proposed by Van Ryzin (1980). This proposal allowed that extrapolation by model could proceed to a risk of 0.01 percent. Later refinements suggested that extrapolation stop at 1 percent to avoid model dependency (Farmer et al., 1982). For an upper bound on linear extrapolation, a straight line from an upper bound on risk at the lower end of the experimental range had been proposed by Gaylor and Kodell (1980).

Judicious choice of a starting point is key to the credibility of a low-dose extrapolation. The high correlations established by Cogliano (1986) and Krewski et al. (1990) led to the idea that ED10s can serve as a common measure for both potency ranking and a starting point for low-dose extrapolation (NRC, 1993). The concerns of Gold et al. (1993), Wartenberg and Gallo (1990), and Hoel (1990) remind the risk assessor that effects at high and low doses can be different, and that the shape of the dose-response curve in the experimental range contains useful information. The concern about overestimation if the dose-response curve turns upward further highlights differences between high and low doses. These concerns can be addressed, in part, by using a point of departure near the lower end of the experimental range.

C. Dose-Response Characterization

Cancer risk assessments attempt to describe a dose-response relationship that can be used to evaluate risks over a range of environmental exposure levels. All relevant experimental information is considered, with incomplete information addressed by a set of default assumptions and adjustments. Some adjustments, for example, the cross-species scaling factor and the time-weighted-average dose metric, are intended as unbiased defaults not expected to contribute to health-conservative estimates (EPA, 1992).

The principal health-conservative defaults in cancer risk assessment have been the presumption that animal results are pertinent to humans, the use of low-dose linear models to extrapolate these results to lower doses, and the use of statistical upper bounds to express slope factors. The proposed cancer guidelines (EPA, 1996a) would not change two of these defaults: animal results would still be considered pertinent to humans, in the absence of convincing evidence to the contrary; and slope factors would retain the use of statistical bounds, though this would be more explicitly communicated by the "L" in the term "LED_{10}."

In contrast, the proposed availability of both linear and nonlinear defaults, each using LED_{10}s as a point of departure, signals a shift from routine application of linear defaults toward a most-plausible-hypothesis approach. When a linear default is used, the LED_{10} approach should have only a small effect on existing slope factors, as linear extrapolation from TD_{50}s or lower points gives results similar to other low-dose linear methods (Krewski, 1990). When a nonlinear default is used, however, there will be substantial changes in the way a cancer hazard is characterized. The risk assessment will characterize how rapidly risk is reduced as exposure is reduced, and this characterization will be both more qualitative and quantitative in nature. Of increased importance will be a discussion of the factors that affect the magnitude of a cancer risk at lower doses, including slope at the point of departure, nature of the response, persistence in the body, sensitivity of humans compared to experimental animals, and nature and extent of human variability in sensitivity. The uncertainty inherent in these factors likely will be much more important than issues in computing LED_{10}s.

Because low-dose linear models have been used extensively in cancer risk assessment, there will be a natural inclination to compare nonlinear methods to linear extrapolation. Even

when linear and nonlinear models both fit the experimental information, risk estimates at lower doses from a nonlinear model can be substantially lower than those from a linear model. This will make it critical that risk assessments using nonlinear methods explain why risk is anticipated to decline more than proportionately with dose, that risk assessments using linear methods explain why risk is anticipated to decline in proportion to dose and, for both methods, identify the major determinants of risk at low doses and address the principal sources of uncertainty and human variability.

VI. FUTURE PLANS

A. Updating of Guidance Document

This guidance document will be updated as new information on the characteristics of modeling of dose-response data and BMD/C analysis becomes available. Information from the literature as well as that submitted to the Agency for review will be considered in revisions to this guidance document. In particular, as approaches become established for selection of the BMR for various endpoints or types of endpoints, this information will be added to the guidance.

The guidance in this document is intended to be consistent with various EPA risk assessment guidelines and with the RfD/C and CRAVE processes (for input into the IRIS Database). As changes occur in either the guidelines or in the RfD/C or CRAVE process, this guidance may also need to be revised.

B. Potential for Use of the BMD/C in Cost-Benefit Analyses

1. Background

As a regulatory agency, the US EPA has responsibility for the implementation in whole or in part of about a dozen environmental statutes. Two statutes explicitly require the Agency to weigh the health and environmental benefits of proposed regulations against their costs: the Federal Insecticide, Fungicide and Rodenticide Act as amended (P.L. 110-532, Oct. 25, 1988, 102 Stat. 2654), and the Toxic Substances Control Act (P.L. 100-551, Oct. 28, 1988, 102 Stat. 2755). President Clinton's Executive Order 12866 on Regulatory Planning and Review also directs the Agency to consider the benefits of the regulations it proposes. In the context of the BMD/C guidance document here, the term "benefits" refers only to the reduced incidence of adverse health effects that would be expected to occur as a result of implementing a given regulation. Currently, no consensus exists as to how noncancer health benefits may be characterized in terms of the expected numbers of cases of some specific disease that are avoided by reducing the level of exposure for a given population. Note that the monetary valuation of these human health benefits and the costing of regulations aimed at obtaining them are outside the scope of risk assessment methodology and are not considered here.

The goal of a benefits analysis is to determine the benefits to society rather than to the individual and thus requires estimates of the incidence of adverse effects in a population, not just unit risk. The estimation of benefits is a multi-step process that requires assessment of individual risk, inter-individual variability and population size in order to estimate population risk. First, the incidence of adverse health effects is estimated for a baseline set of exposure conditions (usually the current condition).

Second, the incidence of expected adverse health effects is estimated for an alternative set of conditions, usually reflecting anticipated reductions in exposure in the affected population after regulations are in effect. Finally, the benefits of achieving the reduced exposure are calculated from the difference between the estimated adverse health effects for the baseline conditions and alternative conditions. While point estimates are often derived for health benefits, an alternate approach could entail a probabalistic scheme, based on upper and lower bounds on the individual risk and exposure curves.

In predicting numbers of expected adverse health effects, one needs to consider the weight of evidence that the critical effect in deriving the BMD/C is the effect of concern in human populations. One would have greatest confidence in benefits assessments based on strong human data; high confidence where the mechanism is known, the critical endpoint is observed in several animal test species and other endpoints occur in the same sequence in several species; and less confidence where a variety of effects are seen in different animal species, appearing in different sequences.

2. Benefits estimates using BMD/C and RfD/C

In the case of carcinogens, standardized methodologies exist for assessment of individual risk and estimates of the incidence of cancer in a population (EPA, 1996a). The use of benchmark-style approaches in the assessment of individual cancer risk has been discussed elsewhere in these guidelines (section V). To estimate benefits, the numbers of expected cases of cancer avoided under different regulatory scenarios are obtained by multiplying individual risk times the size of the exposed population.

In noncancer health benefits analyses, the quantification of the number of cases of disease avoided has been difficult because the RfD/C derivation methodology does not provide estimates

of individual or population risk. To date, with a few exceptions like lead (EPA, 1991b), the primary measure of benefits for regulating non-carcinogens has been the estimated size of the population expected to have reduced exposure to a chemical as the result of regulations implemented to interdict exposure or to reduce ambient concentration of pollutants. These estimates, however, do not directly indicate the extent of morbidity, nor the number of cases of disease or mortality that would be avoided in the affected populations. This guidance does not fill these gaps per se.

The debate regarding the defensibility of approaches to quantifying the benefits associated with reducing exposure to noncarcinogens is long-standing (NRC, 1983). Efforts to quantify noncancer risks at low levels of exposure should be based on empirical dose-response data instead of statistical models to extrapolate to risks at low exposures, and should have a strong biological basis for the particular form of model chosen. In the BMD/C approach, there is an explicit estimate of the dose that corresponds with some level of risk within or not far below the observed range. These estimates potentially can be used to determine the number of cases expected for various non-cancer endpoints when exposure levels are in or near those in the experimental data range.

However, estimation of benefits will usually also require determining risk well below the experimental range, and determining the level below which observed effects are not of concern. These determinations include risks and levels that apply to the general population, and those that apply to sensitive subgroups.

3. Issues in quantifying noncancer health benefits

The benchmark dose approach thus provides a good starting point to develop benefits estimates for non-carcinogens. Many of the key issues for benefits assessment are addressed during BMD/C development, e.g., selecting endpoints relevant to a human disease state, identifying the proper scaling factor to convert animal dose values to human equivalent doses, converting continuous response data to quantal form to support estimates of the probability of expected adverse health effects, and modeling critical health effects data in the observable range using standard mathematical and statistical models aimed at estimating the lower confidence bound on the dose producing a response. Two major issues remain that require the development

of new paradigms to quantify population risk at low dose exposures. One is the extent to which there may be sensitive or "super-exposed" subpopulations. The second issue involves identifying, in a progression of effects in the same mechanistic pathway, those effects which have such limited severity as to be of no concern, beyond representing biomarkers of exposure. Work is continuing in these areas.

C. Research and Implementation Needs

A number of research needs related to the BMD/C approach were indicated in the background document (EPA, 1995c). These included the development of dose-response models and related methods for use with various types of data, guidelines for handling lack of fit, development of methods for applying pharmacokinetic considerations, guidelines for selecting appropriate measures of altered response, study of the sensitivity of the BMD/C to choice of model, particularly in relation to the level of the BMR, and to the confidence limit size, guidelines for selecting a single BMD/C when more than one is calculated, investigation of uncertainty factors, comparison of dose-response curves for different types of data and toxic endpoints, and development of dose-response models for multiple endpoints of toxicity. Several of these research needs have been addressed in the present document. In most cases, however, further research is needed on these topics.

The most critical data need from the point of view of implementing the BMD/C approach on a broad scale according to the guidance in this document is the selection of appropriate measures of altered response. As discussed in Section II.C, selection of the BMR should be dependent on one of two factors, the degree of change considered biologically significant, or that degree of change at the limit of detection. These determinations will need to be done on an endpoint by endpoint basis and, thus, will require bringing together experts in various disciplines to reach consensus. In those situations where biological significance is equated with statistical significance, the limit of detection must be determined. This will require the evaluation of data across a number of studies of similar design to determine for each endpoint of concern what is the limit of detection, on average. Alternatively, studies comparing NOAELs and BMD/Cs for different BMRs could be conducted if an adequate database exists, as has been done for

developmental toxicity (Faustman et al., 1994; Allen et al., 1994a and b; Kavlock et al., 1995).

More experience with modeling data using a variety of models is needed to compare the BMD/Cs obtained and the rationale for selection of the most appropriate model and BMD/C. Sensitivity of the BMD/C to the choice of model, confidence limit size and the actual software used need to be explored with a variety of data sets. It would be advantageous, eventually, to have a small suite of models that could be recommended for noncancer health effects, depending on the type of data.

The development of user-friendly software that is widely available for BMD/C modeling would be extremely helpful. EPA has an on-going effort to develop such software, and the models to be included at the present time are indicated in Appendix E.

Training of risk assessors in the BMD/C approach, application of models and interpretation of data is needed to implement this approach on an Agency-wide basis as well as among individuals outside the Agency. EPA intends to implement such a training program for its employees engaged in risk assessment, as well as risk managers who need to understand the procedure. EPA can also work with other organizations (e.g., professional societies) in developing training programs for those in other sectors who may be involved in risk assessments or serve as reviewers of risk assessments of interest to the EPA.

Several issues have been considered in the development of this guidance document that do not pertain directly to the application of the BMD/C approach but do impinge on its application in the risk assessment arena. These are intentionally not addressed in this guidance document, but do need to be considered further in other efforts. For example, there are several differences in the way in which dose-response analyses are done for cancer endpoints versus noncancer endpoints. These include the assumption of a threshold for noncancer endpoints versus a low-dose linear relationship for cancer (although this is beginning to change with the newly proposed cancer risk assessment guidelines - EPA, 1996). In addition, within human variability is dealt with explicitly in the use of an uncertainty factor for noncancer endpoints, but implicitly for cancer. Dose scaling across species is handled differently for cancer ($mg/kg^{3/4}$-d) versus noncancer (mg/kg/day) health effects. And while linear low-dose extrapolation for cancer to obtain a risk estimate or slope factor includes a risk reduction strategy, there is no implicit risk reduction factor included for

noncancer risk assessment; rather uncertainty factors are used to derive a dose that is considered to be unlikely to cause effects detectable above background. These and other issues, such as how to deal with severity of effect, need further consideration by the Agency.

VII. REFERENCES

Alexeeff, G.V.; Lewis, D.C.; Ragle, N.L. (1993) Estimation of potential health effects from acute exposure to hydrogen fluoride using a 'benchmark dose' approach. Risk Analysis, 13(1):63-69.

Allen, B.C.; Kavlock, R.J.; Kimmel, C.A.; Faustman, E.M. (1994a) Dose-response assessment for developmental toxicity: II. Comparison of generic benchmark dose estimates with NOAELs. Fund. Appl. Toxicol., 23:487-495.

Allen, B.C.; Kavlock, R.J.; Kimmel, C.A.; Faustman, E.M. (1994b) Dose-response assessment for development toxicity: III. Statistical models. Fund. Appl. Toxicol., 23:496-509.

Allen, B.C., and P.L. Strong, C.J. Price, S.A. Hubbard, and G.P. Daston (1996) Benchmark dose analysis of developmental toxicity in rats exposed to boric acid. Fund. Appl. Toxicol. 32: (in press).

Auton, T.R. Calculation of benchmark doses from teratology data. Regulatory Toxicology and Pharmacology, 1994, in press.

Barnes, D.G.; Daston, G.P.; Evans, J.S. Jarabek, A.M.; Kavlock, R.J.; Kimmel, C.A.; Park, C.; Spitzer, H.L. (1995) Benchmark dose workshop: Criteria for use of a benchmark dose to estimate a reference dose. Regulatory Toxicol. Pharmacol., 21:296-306.

Beck, B.D.; Conolly, R.B.; Dourson, M.L.; Guth, D.; Hattis, D.; Kimmel, C.; Lewis, S.C. (1993) Symposium overview: improvements in quantitative noncancer risk assessment. Fund. Appl. Toxicology 20:1-14.

California Office of Environmental Health Hazard Assessment. (1993) Safety assessment for non-cancer endpoints: The benchmark dose and other possible approaches. Summary report.

Catalano, P.J.; Scharfstein, D.O.; Ryan, L.M.; Kimmel, C.A.; Kimmel, G.L. (1993) Statistical model for fetal death fetal weight and malformation in developmental toxicity studies. Teratology 47:281-290.

Chen, C.; Farland, W. (1991) Incorporating cell proliferation in quantitative cancer risk assessment: approaches, issues, and uncertainties. In: Butterworth, B.; Slaga, T.; Farland, W.; McClain, M., eds. Chemical induced cell proliferation: implications for risk assessment. New York: Wiley-Liss, pp. 481–499.

Chen, J.J.; Kodell, R.L.; Howe, R.B.; Gaylor, D.W. (1991) Analysis of trinomial responses from reproductive and developmental toxicity experiments. Biometrics 47:1049-1058.

Clayton, D.; Hills, M. (1993) Statistical Models in Epidemiology. Oxford University Press, Oxford.

Cogliano, V.J. (1986) The U.S. EPA's methodology for adjusting the reportable quantities of potential carcinogens. Proceedings of the 7th National Conference on Management of Uncontrolled Hazardous Wastes (Superfund '86). Washington: Hazardous Materials Control Research Institute, pp. 182–185.

Collins, M.A., G.M. Rusch, F. Sato, P.M. Hext, and R.-J. Millischer. 1995. 1,1,1,2-Tetrafluoroethane: Repeat exposure inhalation toxicity in the rat, developmental toxicity in the rabbit, and genotoxicity in vitro and in vivo. Fund. Appl. Toxicol. 25: 271-280.

Cox, D.R.; Hinkley, D.V. (1974) *Theoretical Statistics*, chapter 7, Chapman and Hall, London.

Crump, K.S. (1984) A new method for determining allowable daily intakes. Fundamental and Applied Toxicology 4:854-871.

Crump, K. S. (1995) Calculation of benchmark doses from continuous data. Risk Analysis 15: 79-89.

Crump, K. S.; R. Howe. (1985) Chapter 9. in *Toxicological Risk Assessment.* D.B. Clayson, D. Krewski, I. Munro, eds. Boca Raton: CRC Press, Inc.

Davidian, M.; Giltinan, D. M. (1995) Nonlinear Models for Repeated Measurement Data. Chapman and Hall, London. Farmer, J.H.; Kodell, R.L.; Gaylor, D.W. (1982) Estimation and extrapolation of tumor probabilities from a mouse bioassay with survival/sacrifice components. Risk Analysis 2(1):27–34.

Dourson, M.L., Hertzberg, R.C., Hartung, R., Blackburn, K. (1985) Novel methods for the estimation of acceptable daily intake. Toxicology and Industrial Health 1:23-41.

Draper, N.; Smith, H. (1981) Applied Regression Analysis, Second Edition. Chapter 10. Wiley, New York.

Farmer, J.H.; Kodell, R.L.; Gaylor, D.W. (1982) Estimation and extrapolation of tumor probabilities from a mouse bioassay with survival/sacrifice components. Risk Analysis 2(1):27–34.

Faustman, E.M.; Allen, B.C.; Kavlock, R.J.; Kimmel, C.A. (1994) Dose-response assessment for developmental toxicity: I. Characterization of data base and determination of NOAELs. Fund. Appl. Toxicol., 23:478-486.

Fleiss, J.L. (1981) Statistical Methods for Rates and Proportions. Second Edition. Wiley. New York.

Gaylor, D.W. (1983) The use of safety factors for controlling risk. Journal of Toxicology and Environmental Health 11:329-336.

Gaylor, D.W.; Kodell, R.L. (1980) Linear interpolation algorithm for low dose risk assessment of toxic substances. J. Environ. Pathol. Toxicol. 4:305–312.

Gaylor, D.; Slikker, W., Jr. (1990) Risk assessment for neurotoxic effects. NeuroToxicology 11: 211-218.

Gaylor, D.W.; Kodell, R.L.; Chen, J.J.; Springer, J.A.; Lorentzen, R.J.; Scheuplein, R.J. (1994) Point estimates of cancer risk at low doses. Risk Analysis 14(5):843–850.

Gerrity, T.R.; Henry, C.J., eds. (1990) Summary report of the workshops on principles of route-to-route extrapolation for risk assessment. In: Principles of route-to-route extrapolation for risk assessment, proceedings of the workshops; March and July; Hilton Head, SC and Durham, NC. New York, NY: Elsevier Science Publishing Co., Inc.; pp. 1-12.

Gold, L.S.; Sawyer, C.B.; Magaw, R.; Backman, G.M.; de Veciana, M.; Levinson, R.; Hooper, N.K.; Havender, W.R.; Bernstein, L.; Peto, R.; Pike, M.C.; Ames, B.N. (1984) A carcinogenic potency database of the standardized results of animal bioassays. Environ. Health Perspect. 58:9–319.

Gold, L.S.; Bernstein, L.; Kaldor, J.; Backman, G.; Hoel, D. (1986a) An empirical comparison of methods used to estimate carcinogenic potency in long-term animal bioassays: lifetable vs summary incidence data. Fund. Appl. Toxicol. 6:263–269.

Gold, L.S.; de Veciana, M.; Backman, G.M.; Magaw, R.; Lopipero, P.; Smith, M.; Blumenthal, M.; Levinson, R.; Bernstein, L.; Ames, B.N. (1986b) Chronological supplement to the carcinogenic potency database: standardized results of animal bioassays published through December 1982. Environ. Health Perspect. 67:161–200.

Gold, L.S.; Slone, T.H.; Backman, G.M.; Magaw, R.; Da Costa, M.; Lopipero, P.; Blumenthal, M.; Ames, B.N. (1987) Second chronological supplement to the carcinogenic potency database: standardized results of animal bioassays published through December 1984 and by the National Toxicology Program through May 1986. Environ. Health Perspect. 74:237–329.

Gold, L.S.; Slone, T.H.; Backman, G.M.; Eisenberg, S.; Da Costa, M.; Wong, M.; Manley, N.B.; Rohrbach, L.; Ames, B.N. (1990) Third chronological supplement to the carcinogenic potency database: standardized results of animal bioassays published through December 1986 and by the National Toxicology Program through June 1987. Environ. Health Perspect. 84:215–286.

Gold, L.S.; Manley, N.B.; Slone, T.H.; Garfinkel, G.B.; Rohrbach, L.; Ames, B.N. (1993) The fifth plot of the carcinogenic potency database: results of animal bioassays published in the general

literature through 1988 and by the National Toxicology Program through 1989. Environ. Health Perspect. 100:65–168.

Gold, L.S.; Manley, N.B.; Slone, T.H.; Garfinkel, G.B.; Ames, B.N. Rohrbach, L.; Stern, B.R.; Chow, K. (in press) Sixth plot of the carcinogenic potency database: results of animal bioassays published in the general literature 1989–1990 and by the National Toxicology Program 1990–1993. Environ. Health Perspect.

Guth (19__)

Haas, C.N. (1994) Dose-response analysis using spreadsheets. Risk Analysis 14(6):1097–1100.

Hasselblad, V.; A.M. Jarabek (1995) Dose-response analysis of toxic chemicals. In: Bayesian biostatistics. D.A. Berry, D.K. Stangl, eds. Marcel Dekker, Inc. New York.

Heindel, J.J., C.J. Price, E.A. Field, M.C. Marr, C.B. Myers, R.E. Morrissey, and B.A. Schwetz (1992). Developmental toxicity of boric acid in mice and rats. Fund. Appl. Toxicol. 18:266-.

Hertzberg, R.C. (1989) Fitting a model to categorical response data with application to species extrapolation of toxicity. Health Physics 57:405-409.

Hertzberg, R.C., Miller, M. (1985) A statistical model for species extrapolation using categorical response data. Toxicology and Industrial Health 1:43-57.

Hext, P.M. and R.J. Parr-Dobrzanski, 1993. HFC 134a: 2 Year inhalation toxicity study in the rat. ICI Central Toxicology Laboratory, Alderley Park, Macclesfield, Cheshire, UK. Report No. CTL/P/3317.

Hoel (1979)

Hoel, D.G. (1990) Assumptions of the HERP index. Risk Analysis 10(4):623–624.

Hoover, S.M.; Zeise, L.; Pease, W.S.; Lee, L.E.; Hennig, M.P.; Weiss, L.B.; Cranor, C. (1995) Improving the regulation of carcinogens by expediting cancer potency estimation. Risk Analysis 15(2):267–280.

Howe, R.B.; Crump, K.S.; Van Landingham, C. (1986) Global 86: a computer program to extrapolate quantal animal toxicity data to low doses. Prepared for U.S. EPA under contract 68–01–6826.

Jarabak, A.M., Hasselblad, V. (1992) Application of a Bayesian statistical approach to response analysis of noncancer toxic effects. Toxicologist 12:98.

Johnson B.L., J. Boyd, J.R. Burg, S.T. Lee, C. Xintaras, and B.E. Albright. 1983. Effects on the peripheral nervous system of workers' exposure to carbon disulfide. Neurotoxicology 4(1): 53-66.

Kavlock, R.J., B.C. Allen, C.A. Kimmel, E.M. Faustman. (1995) Dose-response assessment for developmental toxicity: IV. Benchmark doses for fetal weight changes. Fund. Appl. Toxicol., 26:211-222.

Kavlock, R.J.; Schmid, J.E.; Setzer, R.W., Jr. (1996) A simulation study of the influence of study design on the estimation of benchmark doses for developmental toxicity. Risk Analysis 16:391-403.

Kimmel, C.A., Gaylor, D.W. (1988) Issues in qualitative and quantitative risk analysis for developmental toxicity. Risk Analysis 8:15-20.

Kimmel, C.A.; Wellington, D.G.; Farland, W.; Rose, P.; Manson, J.M.; Chernoff, N.; Young, J.F.; Selevan, S.G.; Kaplan, N.; Chen, C.; Chitlik, L.D.; Siegel-Scott, C.L.; Valaoras, G.; Wells, S. (1989) Overview of a workshop on quantitative models for developmental toxicity risk assessment. Environmental Health Perspectives 79:209-2150.

Kimmel, C.A., M. Siegel, T.M. Crisp, and C.W. Chen (1996) Benchmark concentration (BMC) analysis of 1,3-butadiene (BD) reproductive and developmental effects. Fund. Appl. Toxicol. (Suppl., no. 1, part 2) 30: 146.

Kodell, R.L.; Chen, J.J.; Gaylor, D.W. (1995) Neurotoxicity Modeling for Risk Assessment. Regulatory Toxicology and Pharmacology 22:24-29.

Krewski, D. (1990) Measuring carcinogenic potency. Risk Analysis 10(4):615–617.

Krewski, D.; Zhu, Y. (1994)

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

Krewski, D.; Szyszkowicz, M.; Rosenkranz, H. (1990) Quantitative factors in chemical carcinogenesis: variation in carcinogenic potency. Regul. Toxicol. Pharmacol. 12:13–29.

Krewski, D.; Gaylor, D.; Szyszkowicz, M. (1991) A model-free approach to low-dose extrapolation. Environ. Health Perspect. 90:279–285.

Kupper, L.L.; Hafner, K.B. (1989) How appropriate are popular sample size formulas? The American Statistician 43:101-105.

Lefkopoulou, M.; Moore, D.; Ryan, L. (1989) The analysis of multiple binary outcomes: Application to rodent teratology experiments. Journal of the American Statistical Association 84: 810-815.

Moolgavkar, S.H.; Knudson, A.G. (1981) Mutation and cancer: a model for human carcinogenesis. J. Natl. Cancer Inst. 66:1037–1052.

McCullagh, P.; Nelder, J.A. (1989) Generalized Linear Models, Second Edition. Chapman and Hall, London.

National Research Council (NRC) (1983) Risk Assessment in the Federal Government: Managing the Process. Prepared by: Committee on the Institutional Means for Assessment of Risks to Public Health, Commission on Life Sciences. Washington, DC.

National Research Council (NRC) (1993) Issues in risk assessment. Washington: National Academy Press, pp. 115–116.

National Research Council (NRC) (1994) Science and Judgment in Risk Assessment, Committee on Risk Assessment of Hazardous Air Pollutants, Board on Environmental Studies and Toxicology, Commission on Life Sciences, National Academy Press, Washington, DC.

National Toxicology Program (NTP) (1991) Technical report on the toxicology and carcinogenesis of 1,3-butadiene (CAS No. 106-99-0) in B6C3F$_1$ mice (inhalation studies). U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health, National Toxicology Program. NTP TR434, NIH Publ. No. 92-3165.

Peto, R.; Pike, M.C.; Bernstein, L.; Gold, L.S.; Ames, B.N. (1984) The TD_{50}: a numerical description of the carcinogenic potency of chemicals in chronic-exposure animal experiments. Environ. Health Perspect. 58:1–8.

Price and Berner, 1995. A benchmark dose for carbon disulfide: Analysis of nerve conduction velocity measurements from the NIOSH exposure database. Report to the Chemical Manufacturers Association Carbon Disulfide Panel.

Research Triangle Institute (RTI) (1994) Determination of the no-observable-adverse-effect-level (NOAEL) for developmental toxicity in Sprague-Dawley (CD) rats exposed to boric acid in feed on gestational days 0 to 20, and evaluation of postnatal recovery through postnatal day 21. RTI Identification Number 65C-5657-200.

Ryan, L. 1992. Quantitative risk assessment for developmental toxicity. Biometrics 48:163-174.

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

Simpson (19__).

Setzer, R.W.; Rogers, J.M. (1991) Assessing developmental hazard: the reliability of the A/D ratio. Teratology 44:653-665.

SRA Symposium (1994)

U.S. Environmental Protection Agency (EPA) (1986a) Guidelines for carcinogen risk assessment. Federal Register 51(185):33992–34003.

U.S. Environmental Protection Agency (EPA) (1986b) Science Advisory Board Comments

U.S. Environmental Protection Agency (EPA) (1987) Hazardous substances; reportable quantity adjustments; proposed rules. Federal Register 52(50):8140–8186.

U.S. Environmental Protection Agency (EPA) (1988a) Science Advisory Board Comments.

U.S. Environmental Protection Agency (EPA) (1988b) Science Advisory Board Comments.

U.S. Environmental Protection Agency (EPA) (1988c) Methodology for evaluating potential carcinogenicity in support of reportable quantity adjustments pursuant to CERCLA section 102. Washington: report no. EPA/600/8–89/053.

U.S. Environmental Protection Agency (EPA) (1989a) Reportable quantity adjustments; delisting of ammonium thiosulfate; final rules. Federal Register 54(155):33418–33484.

U.S. Environmental Protection Agency (EPA) (1989b) Technical background document to support rulemaking pursuant to CERCLA section 102, volume 3., Washington: Office of Solid Waste and Emergency Response.

U.S. Environmental Protection Agency (EPA) (1989c) Science Advisory Board Comments.

U.S. Environmental Protection Agency (EPA) (1991a) Guidelines for developmental toxicity risk assessment; notice. Fed Regist, 56:63798-63826.

US Environmental Protection Agency (EPA) (1991b) Regulatory impact analysis of proposed national primary drinking water regulation for lead and copper. Prepared by Wade Miller Associates, Inc. April.

U.S. Environmental Protection Agency (EPA) (1992) Draft report: a cross-species scaling factor for carcinogen risk assessment based on equivalence of $mg/kg^{3/4}/day$; notice. Federal Register 57(109):24152–24173.

U.S. Environmental Protection Agency (1994a) Ranking of pollutants with respect to hazard to human health; proposed rule. Federal Register 59.

U.S. Environmental Protection Agency (1994b) Technical background document to support rulemaking pursuant to the Clean Air Act—section 112(g): ranking of pollutants with respect to hazard to human health. Research Triangle Park, NC: report no. EPA–450/3–92–010.

U.S. Environmental Protection Agency (1994c) Methods for derivation of inhalation reference dose concentrations and application of inhalation dosimetry. Office of Health and Environmental Assessment, Environmental Criteria and Assessment Office, Research Triangle Park, MC. EPA/600/8-90/066F.

U.S. Environmental Protection Agency (1995a) Reportable quantity adjustments; final rule. Federal Register 60(112):30926–30962.

U.S. Environmental Protection Agency (1995b) Technical background document to support rulemaking pursuant to CERCLA section 102, vol. 7. Washington: Office of Solid Waste and Emergency Response.

U.S. Environmental Protection Agency (1995c) The use of the benchmark dose approach in health risk assessment. Office of Research and Development, Washington, DC: EPA/630/R-94/007, February.

U.S. Environmental Protection Agency (1995d) Health assessment document for diesel emissions. Washington, EPA/600/8–90/057Bb.

U.S. Environmental Protection Agency (1995e) Benchmark dose concentration analysis for carbon disulfide. Internal Report.

U.S. Environmental Protection Agency (EPA) (1995f) Proposed guidelines for neurotoxicity risk assessment; notice. Fed Regist, 60:52032-52056

U.S. Environmental Protection Agency (EPA) (1995g) Manganese document

U.S. Environmental Protection Agency (1996a) Proposed guidelines for carcinogen risk assessment. Federal Register 61(79):17960-18011.

U.S. Environmental Protection Agency (1996b) Guidelines for Reproductive Toxicity Risk Assessment; notice. Federal Register (draft).

U.S. Environmental Protection Agency (EPA) (1996c) Integrated Risk Information System (IRIS). Online. National Center for Environmental Assessment, Washington, DC.

Van Ryzin, J. (1980) Quantitative risk assessment. J. Occup. Med. 22(5):321–326.

Wartenberg, D.; Gallo, M.A. (1990) The fallacy of ranking possible carcinogen hazards using the TD_{50}. Risk Analysis 10(4):609–613.

Zeger, S. L.; Liang, K. Y. (1986) Longitudinal data analysis for discrete and continuous outcomes. Biometrics 42: 121-130.

Zhu, Y.; Krewski, D.; Ross, W.H. (1994) Dose-response models for correlated multinomial data from developmental toxicity studies. Applied Statistics 43:583-598.

APPENDIX A

ASPECTS OF DESIGN, DATA REPORTING, AND ROUTE EXTRAPOLATION RELEVANT TO BMD/C ANALYSIS

1. Design

In general, studies with more dose groups and a graded monotonic response with dose will be more useful for BMD/C analysis. Studies with only a single dose showing a response different from controls are not appropriate for BMD/C analysis. Studies in which responses are only at the same level as background or at or near the maximal response level are not considered adequate for BMD/C analysis. It is preferable to have studies with one or more doses near the level of the BMR to give a better estimate of the BMD/C. Studies in which all dose levels show changes compared with control values (i.e., no NOAEL) are readily useable in BMD/C analyses, unless the lowest response level is much higher than that at the BMR.

In a recent simulation study by Kavlock et al. (1996), various aspects of study design (number of dose groups, dose spacing, dose placement, and sample size per dose group) were examined for two endpoints of developmental toxicity (incidence of malformations and reduced fetal weight). Of the designs evaluated, the best results were obtained when two dose levels had response rates above the background level, one of which was near the BMR. In this study, there was virtually no advantage in increasing the sample size from 10 to 20 litters per dose group. When neither of the two dose groups with response rates above the background level was near the BMR, satisfactory results were also obtained, but the BMDs tended to be lower. When only one dose level with a response rate above background was present and near the BMR, reasonable results for the maximum likelihood estimate and BMD were obtained, but here there were benefits of larger dose group sizes. The poorest results were obtained when only a single group with an elevated response rate was present, and the response rate was much greater than the BMR.

2. Aspects of Data Reporting

In most cases, the risk assessor relies on published reports of key toxicological studies in performing a dose-response assessment. Reports from the peer-reviewed literature may contain summary information which can vary in completeness *vis-a-vis* the data requirements of the BMD

method. The optimal situation is to have information on individual subjects. It is very common to have summary information (group level information, e.g., mean and standard deviation) concerning the measured effect, especially for continuous response variables, and it must be determined whether the summary information is adequate for the BMD/C method to proceed.

Dichotomous data are normally reported at the individual level (e.g., 2/10 animals showed the effect). Occasionally a dichotomous endpoint will be reported as being observed in a group, with no mention of the number of animals showing the effect. This usually occurs when the incidence of the endpoint reported is ancillary to the focus of the report. For BMD/C modeling of dichotomous data, both the number showing the response and the total number of subjects in the group are necessary.

Continuous data are reported as a measurement of the effect, such as body weights or enzyme activity in control and exposed groups. The response might be reported in several different ways, including as an actual measurement, or as a contrast (e.g., as absolute change from control or as relative change from control). To model continuous data when individual animal data are not available, the number of subjects, mean of the response variable, and a measure of variability (e.g., standard deviation, SD; standard error, SE; or variance)are needed for each group. The lack of a numerically reported SD/E precludes the calculation of the BMD/C, unless partial information is presented (e.g., SD for control group only) and some assumptions are made. For example, an assumption can be made that the variance in the exposed groups is the same as the controls, but this introduces uncertainty, since the variance in the individual groups allows more precise modeling of the data and calculation of the confidence limits.

Categorical data are defined as a type of quantal data in which there is more than one defined severity category in addition to the no-effect category and the responses in the treatment groups are characterized in terms of the severity of effect (e.g., mild, moderate, or severe histological change). Results may be classified by reporting an entire treatment group in terms of category (group level reporting), or by reporting the number of animals from each group in each category (individual level reporting). For example, a report of epithelial degenerative lesions might state that an exposed group showed a mild effect (group level) or that in the exposed group there were 7 animals with a mild effect and 3 with no effect (individual level reporting). In such a

case, the BMD/C can be calculated using the dichotomous model after combining data in severity categories (e.g., model all animals with an effect or all with greater than a mild effect). Dichotomous data can be viewed as a special case in which there is one category and the possible response is binary (e.g., effect or no effect). Information may also be treated as categorical in cases where an endpoint is inherently a dichotomous or continuous variable, but because the endpoint is reported only descriptively, it cannot be treated quantitatively. The BMD/C approach cannot be applied because the minimum data required for dichotomous models, number affected and total number exposed, are not reported.

Modeling approaches have been discussed for categorical data with multiple categories (Dourson et al., 1985; Hertzberg, 1989; Hertzberg and Miller, 1985) and for group level categorical data (Guth, Simpson). These regression models can also be used to derive a BMD/C, by estimating the probability of effects of severity defined as adverse. This approach is analogous to the BMD/C if the severity categories are defined consistently. This approach has had considerably less review than other approaches being used for the BMD/C.

3. Route Extrapolation

The criteria for determining if route extrapolation is appropriate for risk assessment have been discussed previously (EPA, 1994c; Gerrity and Henry, 1990) and the same criteria apply when selecting data for BMD/C analysis. If it is determined that route extrapolation is appropriate, the general procedure is to convert from the route of exposure in the study to the route of exposure of interest in the risk assessment and then to perform the BMD/C analysis. In this way, any non-linearity in the route extrapolation model would be incorporated into the calculation of the doses used as input into the BMD model.

APPENDIX B

SELECTING THE BENCHMARK RESPONSE (BMR) LEVEL

1. Biologically Significant Change for Specifying the BMR

Continuous endpoints differ from dichotomous endpoints in the way adversity is specified and in the way BMRs can be expressed. Whereas only adverse dichotomous endpoints are selected for consideration in a dose-response assessment, often the adversity of a continuous endpoint depends upon the magnitude of the response. This can be manifested in two general ways. For some continuous endpoints, the NOAEL is defined to be the highest dose at which the difference between treated and control groups does not exceed the criterion for a biologically significant change. In essence, smaller changes are not considered adverse. For example, a decrease in mean adult body weight usually is not considered adverse unless it is at least 10% of the control mean. In such cases, the BMR should correspond to that biologically significant magnitude of change. Selection of the biologically significant magnitude of change is an issue for specialists in various fields of toxicology. Unfortunately, data have seldom been considered in this manner, so that it is difficult to get an answer to the question, "How much of a change is biologically significant and should be considered an adverse effect?" The more usual situation is that the magnitude of change considered biologically significant is based on statistical significance.

Other continuous endpoints are further classified into adverse and non-adverse values by specifying a cut-off value that distinguishes the two categories. For example, human infants weighing less than 2.5 kg at birth have been labeled "low-birth weight," an adverse outcome useful in epidemiological studies of effects of environmental agents on human birth outcomes. Thus, while for dichotomous endpoints, BMRs are expressed in terms of a dose-related increase in the incidence of adverse outcomes, BMRs for continous outcomes may be expressed either in terms of a dose-related change in the mean or, as for dichotomous endpoints, an increase in the incidence over background of the adverse outcome. This latter approach for continuous endpoints mandates a choice of the cutoff distinguishing adverse from non-adverse as well as the choice of the BMR.

2. Limit of Detection for Specifying the BMR

The concept of "limit of detection" for a toxicological bioassay needs to be refined before it can be used to set a specific BMR. Consider a continuous sequence of populations, identical to each other except for the dose of some toxic agent. At least after a threshold has been exceeded, and in the absence of masking effects, the incidence of adverse effects (for dichotomous and dichotomized continuous endpoints) and the mean of continous endpoints will increase (or decrease) as dose increases. As the difference between the response in the control population and in a dosed population increases, so does the probability that a statistical test of that difference, carried out on samples from those populations (as in a toxicological dose-response study), will indicate a significant difference. The probability that the statistical test will be significant when there is a true difference between underlying populations is called the "power" of the test, and depends upon the experimental design. This same concept of power is used to determine sample size when designing studies, when an effort is made to set a sample size that will give a relatively high power (usually 80% - 90%) to detect a given magnitude of effect. Note that, even in the absence of a dose-related effect, a statistical test has some probability of indicating a statistically significant difference. This is variously referred to as the size of the test, or the Type I error, or the "alpha" level of the test (very often, it takes the value of 0.05: i.e., there is a 5% probability of detecting an effect). A way to quantify what is meant by the limit of detection of a bioassay design, then, is by specifying how frequently we would expect to distinguish a treatment group with a response that is <u>just</u> detectable from a control group. Thus, for example, we might decide that a response was at the limit of detection when we could distinguish from background in 50% of the experiments (power of 0.50) in which we attempt it.

To use this concept of limit of detection to specify the BMR for a specific endpoint and species, first a power level is chosen to represent the limit of detection. This should be done in a larger context than just the specific study at hand, and the selection should satisfy the goal for the overall average BMD/C:NOAEL ratio of one. The next step uses the sample size and nesting structure (e.g., litters within dams, repeated measures) for a typical "good[4]" study of the endpoint

---

[4]Using a cadre of standard study designs for a variety of endpoints is one way to reduce the uncertainty that can be introduced when different individuals make decisions from limits of

and species of interest, and historical values of the measurements such as background incidence or control group variance that are necessary for power calculation. With these one finds the magnitude of response just detectable in a two-group design (control and one treatment group) using a one-sided test with a Type I error of 0.05 and the predetermined power (e.g., 0.50). Finally, this change is expressed either as a change in the mean (for continuous endpoints) or as additional or extra risk (for dichotomous endpoints).

Additional risk and extra risk are two ways to quantify the increment to the background risk of an adverse outcome for a dichotomous endpoint. Their definitions are:

$$\text{Additional Risk at dose } d = P(d)-P(0), \text{ and}$$

$$\text{Extra Risk at dose } d = [P(d)-P(0)]/[1-P(0)],$$

where P(d) is the proportion of animals given dose d that have an adverse response.

Additional risk is the proportion of responders in the exposed group beyond that in the control group, and extra risk is the proportion of animals responding that would not otherwise have responded under the assumption that the processes that lead to the adverse outcome in unexposed subjects are independent of the processes that lead to the adverse outcome in the exposed subjects (see Fig. 2). The greater the background incidence, the greater the difference between extra and additional risk. If there are no responders in the control group [P(0)=0], there is no difference between extra and additional risk. For an effect with an incidence of 50% in the control group and 55% in the exposed group, the additional risk is 5% and the extra risk is 10%. Likewise, for a 90% background, and a 1% increase in the exposed group, the additional risk is 1% and the extra risk is 10%. The Agency has used extra risk models in the past for most animal-based

---

detection. We assume here the use of Agency testing protocols as the basis for standard "good" study designs. In areas where standard testing protocols have not been developed, the Agency encourages activities that can assist in identifying a "most common well-designed protocol" for typically-studied endpoints.

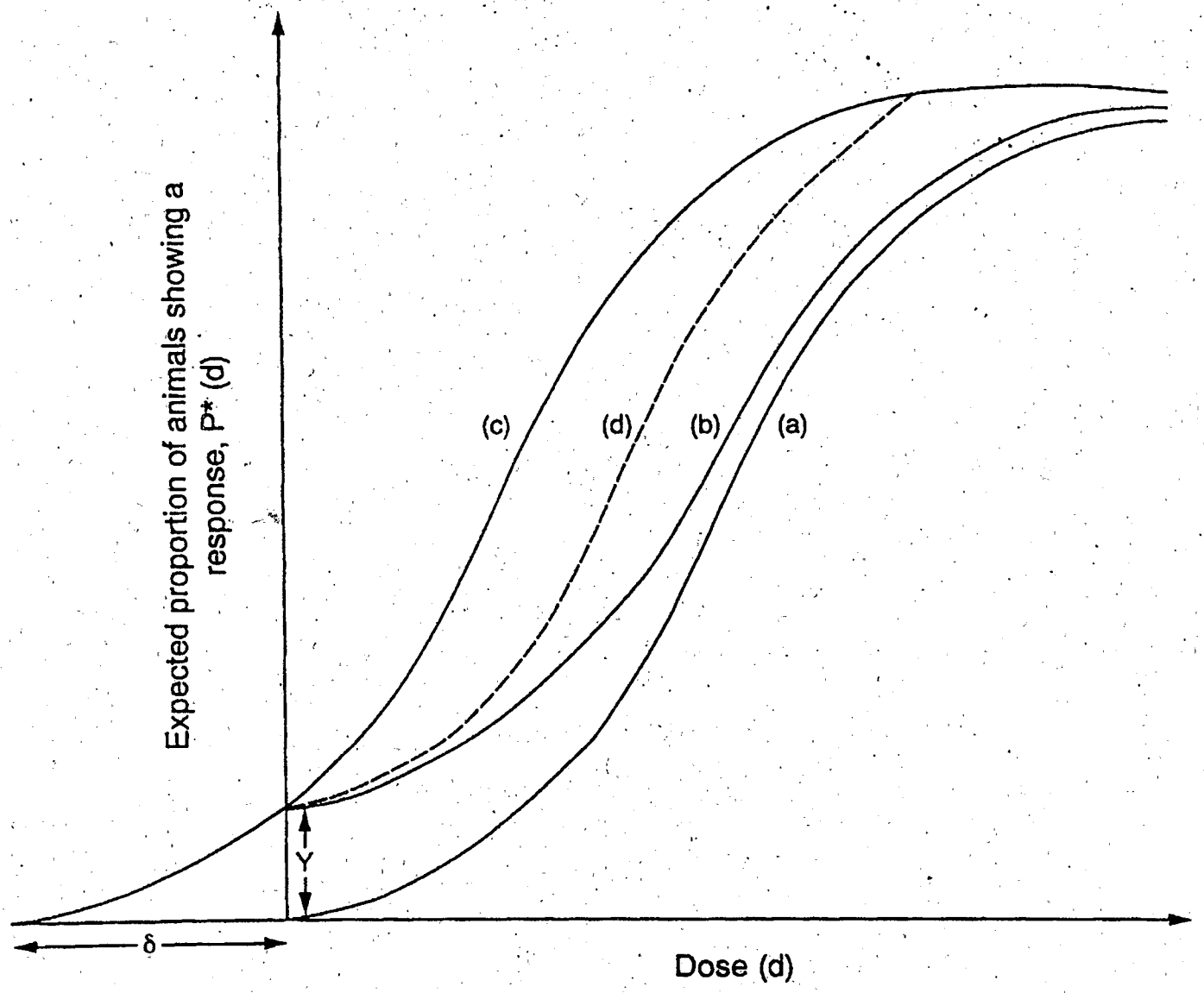


Figure 2. Dose response curves for models incorporating difference forms of spontaneous background response

(a) No background response ($P^*(d) = P(d)$).

(b) Independent background response ($P^*(d) = \gamma + (1 - \gamma) P(d)$) = extra risk.

(c) Additive background response ($P^*(d) = P(d + \delta)$).

(d) Additional background response ($P^*(d) = \gamma + P(d)$).

cancer risk assessments and for most work with BMD/C analyses to date. EPA-supported research on developmental toxicity data (Allen et al., 1994a, b; Kavlock et al., 1995) used additional risk, but since the background incidence in the data used were relatively low, the difference between additional and extra risk are likely to be minimal.

It has been proposed that the extra risk approach is consistent with the assumption that independent mechanisms are responsible for the background incidence and the excess incidence of effect in the exposed group. Although the basis for this explanation has not been well described (Hoel, 1979), it could be argued that the effects of exposure would be masked in proportion to the background response, since independent mechanisms would not be assumed to have an additive effect. The true exposure-related response would then be reflected by the extra risk model. There is no correspondingly simple interpretation of additional risk; its use seems primarily to reflect computational convenience.

There is limited basis for a science policy choice as to whether to use an additional or extra risk approach for BMD/C analyses. The choice of the extra risk model would be conservative in the sense that, for a given increment of risk, say 10%, the BMD/C corresponding to an extra risk of 10% would always be equal to or less than that corresponding to an additional risk of 10%. When the BMR is set based on limit of detection considerations, however, it does not matter which of the two risk formulations are used; the resulting BMD/C will be the same, because the value of the BMR based on extra risk and additional risk that correspond to the same limit of detection differ (see the example, below). The default approach for BMD/C analysis of dichotomous outcomes is to use extra risk to specify BMRs.

3. Examples

Suppose the endpoint of interest has a dichotomous response and no nested structure, so that we would normally assume that the response has a binomial distribution. Fleiss (1981) gives an approximate formula for the power of a two-sample test in such a situation, when each group is the same size. This is a function of sample size, the probability of response in each group, and the Type I error of the test. This formula (Fleiss, 1981) can be used for the task at hand by fixing the incidence of the adverse outcome in the control group, the sample size, and the desired Type I error, and then, either by trial and error or by some more efficient search method, solving for the

treatment group response that would be detectable with the specified power under those conditions. Figure 3 shows the result of doing this for a range of values for control incidence, Type I error of 0.05 in a one-sided test, a sample size of 25 per group, and three powers: 0.2, 0.5, and 0.8. The "barely detectable response" is plotted both as extra risk (top panel) or additional risk (lower panel). Note that, for a given control incidence, the same treatment group incidence results in a higher value for extra risk than it does additional risk, because of the "1 - P(0)" term in the denominator for extra risk.

As an example of how to read this graph, suppose we are looking at data for an endpoint with a background incidence that is usually around 10%, and which is usually measured in a study with about 25 animals per group. Go to the figure that corresponds to the way you want to express risk ("Extra" or"Additional"), find 0.1 on the x-axis (there is a vertical dotted line there), and follow it up until it crosses one of the plotted curves. The vertical line crosses that thinnest, bottommost curve at an additional risk of about 0.13 or an extra risk of about 0.15. Since this curve corresponds to a power of 0.2, this means that one in five (20%) experiments in a situation with a control incidence of 0.1, an additional risk in the treatment group of 0.13 (that is, the treatment response is about 0.23),with 25 animals per group, and a one-sided test with a Type I error of 0.05, would indicate that the treatment response was greater than control. To raise the proportion of experiments that indicate that the treatment response was greater than control to one in two (50%), the additional risk for the treatment group would have to be about 0.23, and to raise it to eight out of ten (a typical pre-design-stage sort of power), the additional risk would have to be about 0.35.

If the magnitude of treatment effect (difference between the mean responses of treatment and control groups) that is deemed biologically significant is the smallest treatment effect that a standard experimental design for that endpoint could detect, the BMR would be based on the limit of resolution of the standard bioassay for looking at that endpoint in the relevant test species. For example, for average litter sizes, the variance of litter means of term fetus weight for CD-1 mice is about 0.008 with a mean of about 1.02 gm (Setzer and Rogers, 1991). A commonly used sample size formula (Kupper and Haffner,1989) can be rearranged to express the difference between two

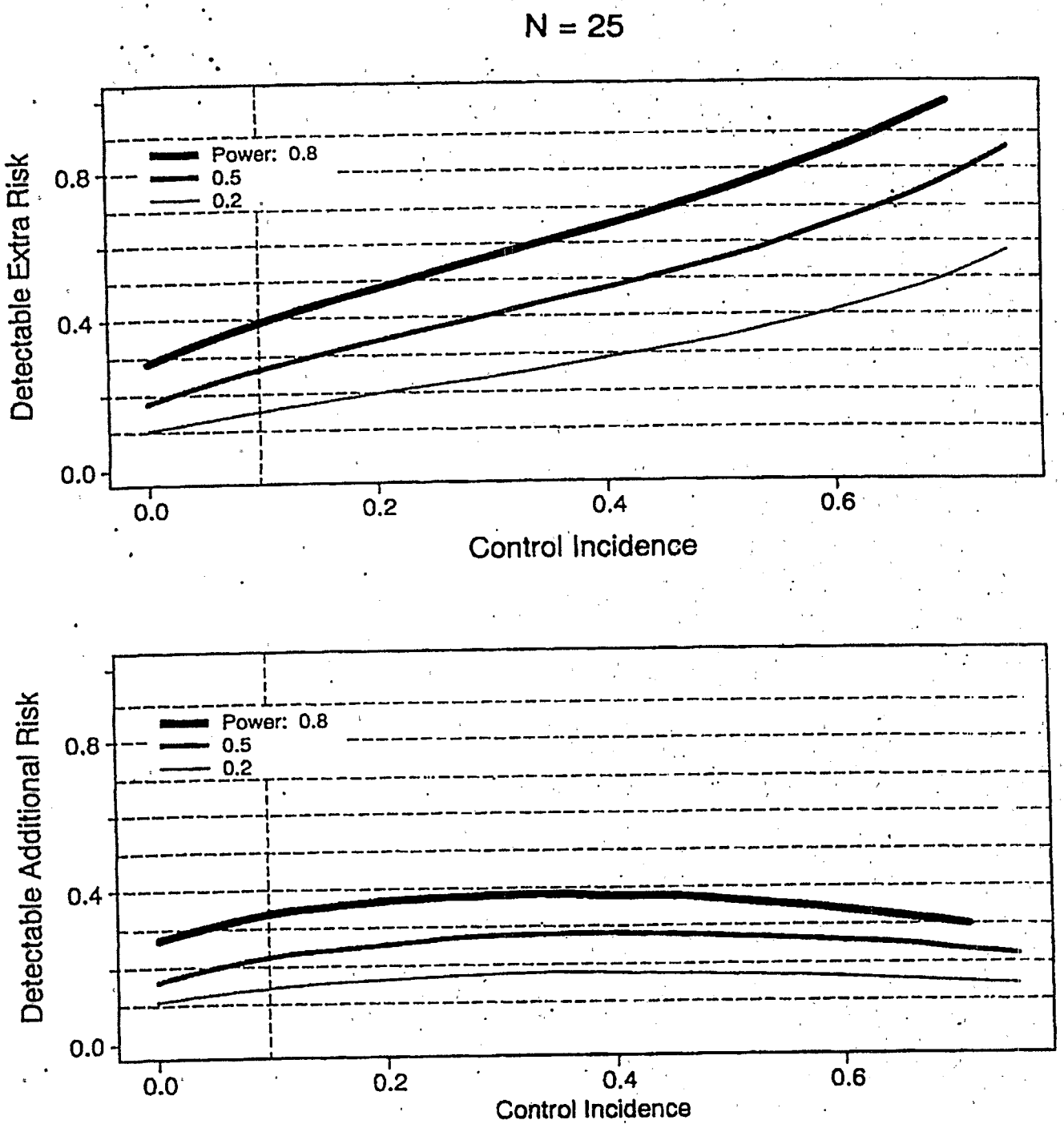


Figure 3. Detectable extra risk or additional risk for a sample size of 25 (N)

means just detectable with a test of a specified type I error and power:

$$\delta = \sigma \frac{(Z_{1-\alpha} + Z_{1-\beta})}{\sqrt{n/2}},$$

where δ is the desired difference between control and treatment group means, σ is the control group standard deviation, n is the number of animals in a group, α is the Type I error, $1 - \beta$ is the desired power, and $Z_{1-\alpha}$ is the number from a table of the normal distribution such that the probability that a standard normal random variable is less than $Z_{1-\alpha}$ is $1-\alpha$.

This gives a detectable difference of about 0.04 grams (about 4%) using a test with a type I error of 0.05 and a power of 0.5 (for powers of 0.2 and 0.8 the values are 0.02 gm and 0.08 gm, respectively). If it were decided that the limit of detectability would be based on a difference detectable with a power of 0.5, then, to reflect the limit of detectability of the conventional assay, the BMR would be set at a difference of 0.04 gm.

4. Selecting the Critical Power Level for the Limit of Detectability

The critical power level to use for the "limit of detectability" with any particular study design needs to be determined in advance of applying this approach to setting the BMR. This could be determined using data from a large number of studies of similar "standard" design as was done for developmental toxicity data (Allen et al., 1994 a, b; Kavlock et al., 1995), or using simulation techniques based on standard study design and background incidence. Other requirements for setting the BMR using this approach are general agreements on the details of the behaviors of those endpoints that affect power calculation; for example, the "typical" study design for evaluating an endpoint in a given species; for dichotomous responses, normal background incidences; and for continuous response, normal control variance levels and changes considered to be "biologically significant" or cutoff levels distinguishing adverse from non-adverse values of the continuous endpoint.

APPENDIX C
MATHEMATICAL MODELING

1. Introduction

Dose-response models for toxicology data are usually of the type called "nonlinear" in mathematical terminology. In a linear model, the value the model predicts is a linear combination of the parameters. For example, in a linear regression of a response y on dose, the predicted value is a linear combination of a and b, namely, $a \times 1 + b \times dose$. Note that, even a quadratic or other polynomial is a linear model, in this sense: $y = a + b \times dose + c \times dose^2 + d \times dose^3$ is a third order polynomial (a cubic) equation, but is still a linear combination of the parameters, a, b, c, and d. In contrast, in a nonlinear model, for example the log-logistic with background,

$$p = P_0 + (1 - P_0) \times \frac{1}{1 + e^{-[a + b \times \ln(dose)]}},$$

the response is not a linear combination of the parameters (here, P_0, a, and b). The distinction is important, because nonlinear models are usually more difficult to fit to data, requiring more complicated calculations, and statistical inference is more typically approximate than with linear models. Note that this definition of "linear" is in contrast to the way the term is used in reference to cancer dose-response assessment, in which the phrase "low-dose linear" refers to models in which the slope is positive at zero dose.

This section will discuss some aspects of model-fitting: initial model selection, approaches to fitting models to data, ways to select among several models fitted to the same data set, and calculation of the confidence intervals..

2. Model Selection

The initial selection of a group of models to fit to the data is governed by the nature of the measurement that represents the endpoint of interest and the experimental design used to generate the data. In addition, certain constraints on the models or their parameter values sometimes need to be observed, and may influence model selection. Finally, it may be desirable to model multiple

endpoints at the same time. The diversity of possible non-cancer endpoints and shapes of their dose-responses for different toxicants precludes specifying a small set of models for further consideration. This will inevitably lead to the use of judgement and occasional ambiguity when selecting a final model to calculate a BMD/C. It is hoped that, as experience using benchmark dose methodology in dose-response assessment accumulates, it will be possible to narrow the number of acceptable models.

a. Type of endpoint

The kind of measurement variable that represents the endpoint of interest is an important consideration in selecting mathematical models. Commonly, such variables are either continuous, like liver weight or the specific activity of a given liver enzyme, or discrete, commonly dichotomous, like the presence or absence of abnormal liver status. However, other types are common in biological data; for example: ordered categorical, like a histology score that ranges from 1-normal to 5-extremely abnormal; counts, such as counts of deaths or the numbers of cases of illness per thousand person-years of exposure to a given exposure condition; waiting time, such as the time it takes for an illness to appear after exposure, or age at death, or multiple endpoints considered jointly (see, for example, Krewski and Zhu, 1995, Lefkopoulou et al. 1989). It is beyond the scope of this document to consider all possible kinds of variables that might be encountered, so further discussion will concentrate on dichotomous and continuous variables.

Dichotomous variables. Data on dichotomous variables are commonly presented as a fraction or percent of individuals that present with the given condition at a given dose or exposure level. For such endpoints, normally we select probability density models like logistic, probit, Weibull, and so forth, whose predictions lie between zero and one for any possible dose, including zero. The natural form of some models, such as the log-logistic model, presumes that the proportion of controls with the abnormal response (an estimate of background) will always be zero. The default approach taken in this document is not to include a background term in the model unless there is evidence of a background response, or if the first 2 or more dosed groups do not show a monotonic increase in response. This is an area where more work is needed.

Continuous variables. Data for continuous variables are often presented as means and standard deviations or standard errors, but may also be presented as a percent of control or some other

standard. From a modeling standpoint, the most desirable form for such data is by individual. Unlike the usual situation for dichotomous variables, summarization of continuous variables results in a loss of information about the distribution of those variables.

The preferred approach to expressing the BMR will determine the approach to modeling continuous data. Two broad categories of approach have been proposed: 1) to express the BMR as a particular change in the mean response, possibly as a fraction of the control mean, a fraction of the standard deviation of the measurement from untreated individuals, or a level of the response that expert opinion holds is adverse; or 2) to decide on a level of the outcome to consider adverse, and treat the proportion of individuals with the adverse outcome much as one would a dichotomous variable (see Appendix B above).

Typical models to use in the first situation include linear and polynomial models, and power models or other nonlinear models. In the second situation, one approach is to classify each individual as affected or not, and model the resulting variable as dichotomous. An alternative is to use a hybrid approach, such as that described by Gaylor and Slikker (1990), Kodell et al. (1995), and Crump (1995), which fit continuous models to continuous data, and, presuming a distribution of the data, calculate a BMD/C in terms of the fraction affected.

b. Experimental design

The aspects of experimental design that bear on model selection include the total number of dose groups used and possible clustering of experimental subjects. The number of dose groups has a bearing on the number of parameters that can be estimated: the number of parameters that affect the overall shape of the dose-response curve generally cannot exceed the number of dose groups.

Clustering of experimental subjects is actually more of an issue for methods of fitting the models than for choice of the model form itself. The most common situation in which clustering occurs is in developmental toxicology experiments, in which the toxicant is applied to the mother, and individual offspring are examined for adverse effects. Another example is for designs in which individuals yield multiple observations (repeated measures). This can happen, for example, when each subject receives both treatment and control (common in studies with human subjects), or each subject is observed multiple times after treatment (e.g., neurotoxicity studies). The issue

in all these examples is that individual observations cannot be taken as independent of each other. Most methods used for fitting models rely heavily on the assumption that the data are independent, and special fitting methods need to be used for data sets that exhibit more complicated patterns of dependence (see, for example, Ryan 1992; Davidian and Giltinan, 1995).

c. Constraints and covariates

An obvious constraint on models for dichotomous data has already been been discussed: probabilities are constrained to be positive numbers no greater than one. However, biological reality may impose other constraints on models. For example, most biological quantities are constrained to be positive, so models should be selected so that their predicted values, at least in the region of application, conform to that constraint. In models in which dose is raised to a power which is a parameter to be estimated (such as a Weibull model), if that parameter is allowed to be less than 1.0, the slope of the dose-response curve becomes infinite at a dose of zero. This is often seen as an undesirable situation, and the default is to constrain these parameters to be greater than or equal to 1.

It is sometimes desirable to include covariates on individuals when fitting dose-response models. For example, litter size has often been included as a covariate in modeling laboratory animal data in developmental toxicity. Another example is in modeling epidemiology data, when certain covariates (e.g., age, parity) are included that are expected to affect the outcome and might be correlated with exposure. In continuous models, if the covariate has an effect on the response, including it in a model may improve the precision of the overall estimate by accounting for variation that would otherwise end up in the residual variance. In any kind of model, any variable that is correlated with dose, and which affects outcome, would need to be included as a covariate.

3. Model Fitting

The goal of the fitting process is to find values for all the model parameters so that the resulting fitted model describes those data as well as possible; this is termed "parameter estimation." In practice, this happens when the dose-group means predicted by the model come as close as possible to the data means. One way to achieve this is to write down a function of all the parameters (the objective function) and all the data, with the property that the parameter

values that correspond either to an overall minimum (or, equivalently, an overall maximum) of the function, or that result in function values of zero, give the desired model predictions. The actual fitting process is carried out iteratively. Many models will converge to the right estimates for most data sets from just about any reasonable set of initial parameter values; however, some models, and some data sets, may require multiple guesses at initial values before the model converges. It also happens occasionally that the fitting procedure will converge to different estimates from different initial guesses. Only one of these sets of estimates will be "best". It is always good practice when fitting nonlinear models to try different initial values, just in case.

There are a few common ways to construct objective functions (estimates): the methods of nonlinear least squares, maximum likelihood, and generalized estimating equations (GEE). The choice of objective function is determined in large part by the nature of the variability of the data around the fitted model. The method of nonlinear least squares, where the objective function is the sum of the squared differences between the observed data values and the model-predicted values, is a common method for continuous variables when observations can be taken as independent. A basic assumption of this method is that the variance of individual observations around the dose-group means is a constant across doses. When this assumption is violated (commonly, when the variance of a continuous variable changes as a function of the mean, often proportional to the square of the mean, giving a constant coefficient of variation), a modification of the method may be used in which each term in the sum of squares is weighted by an estimate of the variance at the corresponding dose. This method is especially appropriate when the data to be fitted can be supposed to be at least approximately normally distributed.

Maximum likelihood is a general way of deriving an objective function when a reasonable supposition about the distribution of the data can be made. Because estimates derived by maximum likelihood methods have good statistical properties, such as asymptotic normality, maximum likelihood is often a preferred form of estimation when that assumption is reasonably close to the truth. An example of such a situation is the case of individual independently treated animals (e.g., not clustered in litters) scored for a dichotomous response. Here it is reasonable to suppose that the number of responding animals follows a binomial distribution with the probability of response expressed as a function of dose. Continuous variables, especially means of several

observations, are often normal (gaussian) or log-normal. When variables are normally distributed with a constant variance, minimizing the sum of squares is equivalent to maximizing the likelihood, which explains in part why least squares methods are often used for continuous variables. In developmental toxicity data, the distribution of the number of animals with an adverse outcome is often taken to be approximately beta-binomial. This particular likelihood is used to accomodate the lack of independence among littermates.

A third group of approaches to estimating parameters are the related quasi-likelihood method (McCullagh and Nelder, 1989) and the method of GEE (see Zeger and Liang, 1986), which require only that the mean, variance, and correlation structure of the data be specified. GEE methods are similar to maximum likelihood estimation proceedures in that they require an iterative solution, provide estimates of standard errors and correlations of the parameter estimates, and estimates are asymptotically normal. Their use so far has primarily been to handle forms of lack of independence, as in litter data, and would be useful in any of a number of kinds of repeated measures designs, such as occur in clinical studies and repeated neurobehavioral testing.

4. Assessing How Well the Model Describes the Data

An important criterion is that the selected model should describe the data, especially in the region of the BMR. Most fitting methods will provide a global goodness-of-fit measure, usually associated with a P-value. These measures quantify the degree to which the dose-group means that are predicted by the model differ from the actual dose-group means, relative to how much variation of the dose-group means one might expect. Small P-values (say, $P < 0.05$) indicate that it would be unlikely to achieve a value of the goodness-of-fit statistic at least this extreme if the data were actually sampled from the model, and, consequently, the model is a poor fit to the data. Larger values cannot be compared from one model to another since they assume the different models are correct; they can only identify those models that are consistent with the experimental results. When there are other covariates in the models, such as litter size, the idea is the same, just more complicated to calculate. In this case, the range of doses and other covariates is broken up into cells, and the number of observations that fall into each cell is compared to that predicted by the model.

It can happen that the model is never very far from the data points (so the P-value for the

goodness-of-fit statistic is not too small), but is always on one side or the other of the dose-group means. Also, there could be a wide range in the response, and the model predicts the high responses well, but misses the low dose responses. In such cases, the goodness-of-fit statistic might not be significant, but the fit should be treated with caution.

The best way to detect the form of these deviations from fit is with graphical displays. Plots should always supplement goodness-of-fit testing. It is extremely helpful that plots that include data points also include a measure of dispersion of those data points, such as confidence limits for the means, if the data points are means.

In certain cases, the typical models for a standard study design cannot be used with the observed data as, for example, when the data are monotonic, or when the response rises abruptly after some lower doses that give only the background response. In these cases, adjustments to the data (e.g., a log-transformation of dose) or the model (e.g., adjustments for unrelated deaths) may be necessary. In the absence of a mechanistic understanding of the biological response to a toxic agent, evidence from exposures that give responses much more extreme than the BMR does not really tell us very much about the shape of the response in the region of the BMR. Such exposures, however, may very well have a strong effect on the shape of the fitted model in the region of the BMD/C. Thus, if lack of fit is due to characteristics of the dose-response data for high doses, the data may be adjusted by eliminating the high dose group. The practice carries with it the loss of a degree of freedom, but may be useful in cases where the response plateaus or drops off at high doses. Alternatively, an entirely different model could be fit.

5. Comparing Models

It will often happen that several models provide an adequate fit to a given data set. These models may be essentially unrelated to each other (for example a logistic model and a probit model often do about as well at fitting dichotomous data) or they may be related to each other in the sense that they are members of the same family that differ in which parameters are fixed at some default value. For example, one can consider the log-logistic, the log-logistic with non-zero background, and the log-logistic with threshold and non-zero background to all be members of the same family of models. Goodness-of fit statistics are not designed to compare different models, so alternative approaches to selecting a model to use for BMD/C computation need to be

pursued.

When other data sets for similar endpoints exist, an external consideration can be applied. It may be possible to compare the result of BMD/C computations across studies of all the data that were fit using the same form of model, presuming that a model can be found that describes all the data sets. Another consideration is the existence of a conventional approach to fitting a kind of data. In this case, communication with specialists in that data is eased when a familiar model is used to fit the data. Neither of these considerations should be seen as justification for using ill-fitting models. Finally, it is generally considered preferable to use models with fewer parameters, when possible.

Generally, both in the method of least squares and in maximum likelihood methods, within a family of models, as additional parameters are introduced the objective function will appear to improve in behavior. This general behavior is due solely to the increase in the additional parameters. Likelihood ratio tests can be used to evaluate whether the improvement in fit afforded by estimating additional parameters is justified. Such tests cannot be applied to compare models from different families, however. Some statistics, notably Akaike's Information Coefficient (AIC) can be used to compare models with different numbers of parameters using a similar fitting method (for example, least squares or a binomial maximum likelihood). Although such methods are not exact, they can provide useful guidance in model selection.

6. Using Confidence Limits to Get a BMD/C

When a BMD/C is determined, it requires the choice of a model together with a suitable method to calculate a confidence limit, as well as the choice (or prior specification) of the level of confidence and the magnitude of the endpoint. Confidence limits bracket those models which, within a particular model family, are consistent with the data. They do not make statements about the extent to which a model is likely to be the true one (see Cox and Hinkley, 1974).

The choice of endpoint or its severity is discussed elsewhere in this document. The "level" at which confidence limits are set has tended to be 95% for a variety of applications (e.g., prediction of future monitoring values, human cancer prediction). This level reflects the extent to which outlying values are tolerable for the system and is chosen to cover some reasonable amount of the distribution of the source of the modeled data. This cannot account for or assume any

correspondence between the modeled animal data and the human population of concern. Rather, the "confidence" associated with an interval indicates the percent of repeated intervals based on experiments of the same sort that are expected to cover (include) the dose associated with the BMR. The choice of confidence level represents tradeoffs in data collection costs and the needed data precision, just as hypothesis-testing levels do. Just as 0.05 is a convenient (but not necessarily good for all data) level for tests, 95% is a convenient choice for most limits. An example of the effects of choice among 90, 95, and 99% for two different models may be found in the background document (EPA, 1995c).

The method by which the confidence limit is obtained is typically related to the manner in which the BMD/C is estimated from the model. Historically, most Agency modelling has been likelihood based (see discussion below); mostly this reflects general trends in modelling. How well any method behaves (e.g., how narrow the confidence interval of fixed level is) relative to other methods with the same model can depend on how nearly the effect level for estimation is contained in the exposure range in the source study. Different types of models lend themselves to different types of estimation, depending on such things as the extent to which the model accommodates clustered data and whether the model equation can be restated in terms of the dose that is calculated to produce the specific effect. Software used to determine benchmark doses should identify the estimation procedure used and the available confidence limits.

Several ways to derive model functions are described earlier in this Appendix. For both likelihood and estimating equation strategies, the most typical approach to constructing confidence limits relies on asymptotic normality to establish confidence sets, although neither strategy requires that the actual endpoint of interest be a continuous variable.

With regard to likelihood-based methods, confidence intervals (CIs) based on the asymptotic distribution of the likelihood ratio are preferred to those based on the asymptotic distribution of the MLEs, because they can use a commonly tabled distribution function, but both can give problems in ranges where the assumptions needed to use asymptotic theory begin to weaken (e.g., as sample sizes decrease, as interest focuses farther from the experimental doses, as observations become more correlated). In general, however, it is preferred to base CIs for parameters estimated by maximum likelihood across various data contexts on the asymptotic

distribution of the likelihood ratio, owing to their tendency to give better coverage behavior[5] (Crump and Howe, 1985).

In the long run, there may be a preference for one type of construction with GEE (e.g., like Ryan, 1992, who uses the delta method and sample estimates). It still has to be demonstrated whether those types are similar in performance (e.g., coverage behavior) for their respective parameters (where GEE were used) to those from the asymptotic distribution of the likelihood ratio in the likelihood context.

Thus, the BMD/C is determined by 1) selecting an endpoint, 2) identifying a BMR (a predetermined level of change in response relative to controls), 3) establishing, by an appropriate estimation procedure, a model that fits the data adequately, and 4) calculating a confidence limit at the BMR using the model and the same estimation procedure.

At the time of this draft, some commercial software is available that is designed specifically for carrying out steps 3) and 4) by maximum likelihood methods. EPA is currently developing software for this purpose that will be made widely-available to all potential users. GEE solutions, which might be used with nested data, involve iterative fitting after the method of weighted least squares but standardized routines are available for certain contexts; these are written as commonly-used software macros (e.g., in SAS) that must be inserted into user-fashioned programs.

---

[5]While we say there is, for instance, a 95% probability that our interval covers the BMR, the actual probability may be something else, sometimes even as low as 30% or 40%, although more usually in the 50% to 90% range. This can happen when we have not allowed for something like correlated observations in the model. Then it is said that the nominal and actual coverages diverge. It is not clear if anyone has looked at the extent of this divergence for various construction approaches for GEE CIs.

APPENDIX D

EXAMPLES OF BMD/C ANALYSES

Example #1: Carbon Disulfide

Summary of Main Points Illustrated

Determination of the BMR for a continuous variable based on biological significance of response.

Use of individual rather than group exposure and response data.

Summary of Data

Study: Occupational inhalation study (in humans) by NIOSH.
Endpoints modeled: Sural and peroneal nerve conduction velocity and amplitude
Endpoint used: peroneal nerve conduction velocity

The benchmark concentration for carbon disulfide was derived from nerve conduction velocity changes reported in the occupational study by Johnson et al., 1983. Johnson et al. reported several neurophysiological measurements in workers exposed for an average of 12 years. Endpoints affected by exposure included peroneal nerve conduction velocity and amplitude, and sural nerve conduction velocity. The published report presented exposure groups categorized as low, medium and high exposure and presented mean and median exposures for each group (Table 1).

Table 1. Group data used in the Benchmark Dose Calculation for Carbon Disulfide - As summarized in Johnson et al., 1983.

	Exposure Level (ppm)			Peroneal NCV		
Group	Mean	SD	Median	N	Mean	SD
Comparison group			0.2	196	45.3	4.4
Low	1.2	1.0	1.0	44	43.7	5.1
Medium	5.1	4.1	4.1	56	43.4	4.8
High	12.6	26.9	7.6	36	41.8	4.5

The Chemical Manufacturers Association Carbon Disulfide Panel obtained the raw individual data from the Johnson study from NIOSH and performed a benchmark concentration analysis which is presented in the report by Price and Berner, 1995.

Selection of the BMR Level

Benchmark response: 10% change (note that this BMR is > the response in the highest exposure group)

The decision of the RfC work group was to use a 10% decrease in nerve conduction velocity as the definition of the BMC. The BMR is therefore at a higher level of response than the highest dose group in the study and was also higher than a statistically significant response level in the study. This decision was aided and supported by extensive discussion with neurotoxicologists from EPA's NHEERL, who attended the work group meeting. There was considerable discussion about the proper choice of the BMR because use of the BMC10 resulted in a BMC that was higher than any of the exposure group responses, i.e., the BMC involved extrapolation above the data points. This approach was supported on the grounds that the response observed in the high concentration group was considered to be, at most, mildly adverse (a peer-reviewer's comment was that the effects in the Johnson study should be considered pre-adverse), because a change in conduction velocity of 10% is about where a clinician would begin to be concerned, and because a 10% change is about equal to one standard deviation for this endpoint. The use of the individual subject analysis was preferred because it allowed for age adjustment.

Calculation of the Benchmark Concentration

Model: polynomial model used, data fit with linear term only

A. Group Level Data

Analyses were done initially using the group level data based on the arithmetic mean exposures (with no variability estimate) using commercial software (Table 2). These results are presented in the internal EPA report 'Benchmark Concentration Analysis for Carbon Disulfide' (EPA, 1995e). (Also presented in this report are BMC analyses on the developmental studies by Tabacova and colleagues.)

B. Individual Data

The Chemical Manufacturers Association Carbon Disulfide Panel performed a benchmark concentration analysis on the individual data, which allowed for an age adjustment to be included in the analysis and allowed for exposure variability to be considered because each subject had a unique exposure and effect measurement (Price and Berner, 1995).

Further analysis of the raw data by EPA involved correction of some missing or implausible data points and evaluation of the interaction of age and exposure. This analysis found

a greater decline in conduction velocity with age in exposed compared with control subjects, suggesting an interaction of age and exposure, and found cumulative exposure to better explain the response (see figure). The resulting BMCs are shown in Table 2.

Table 2. Peroneal Nerve Conduction Velocity Benchmark Concentration Results

	BMR	BMC ppm	BMC mg/m3
Group Data	10%	12	37
Individual Data	10%	20	62

The difference in the BMCs between the analysis based on group data versus individual data are due mainly to the skewed distribution of concentrations in the defined exposure categories for the group analysis (See Table 1).

Explanation of Figure 4:

The top horizontal solid line is the mean adjusted nerve conduction velocity (NCV) for the '"typical" subject (34 year old, 70 inches tall, 177 pound). The horizontal dashed line and lower solid line show the BMR of 5 or 10 % reduction in peroneal NCV. The dashed sloping line shows the 95% confidence limit on the linear model using the cumulative exposure. The vertical line connects the intersection of the lower confidence limit and the BMR of a 10% reduction to the x-axis, showing the BMC. The solid sloping line shows the lower confidence limit on the regression model using mean exposure index for comparison.

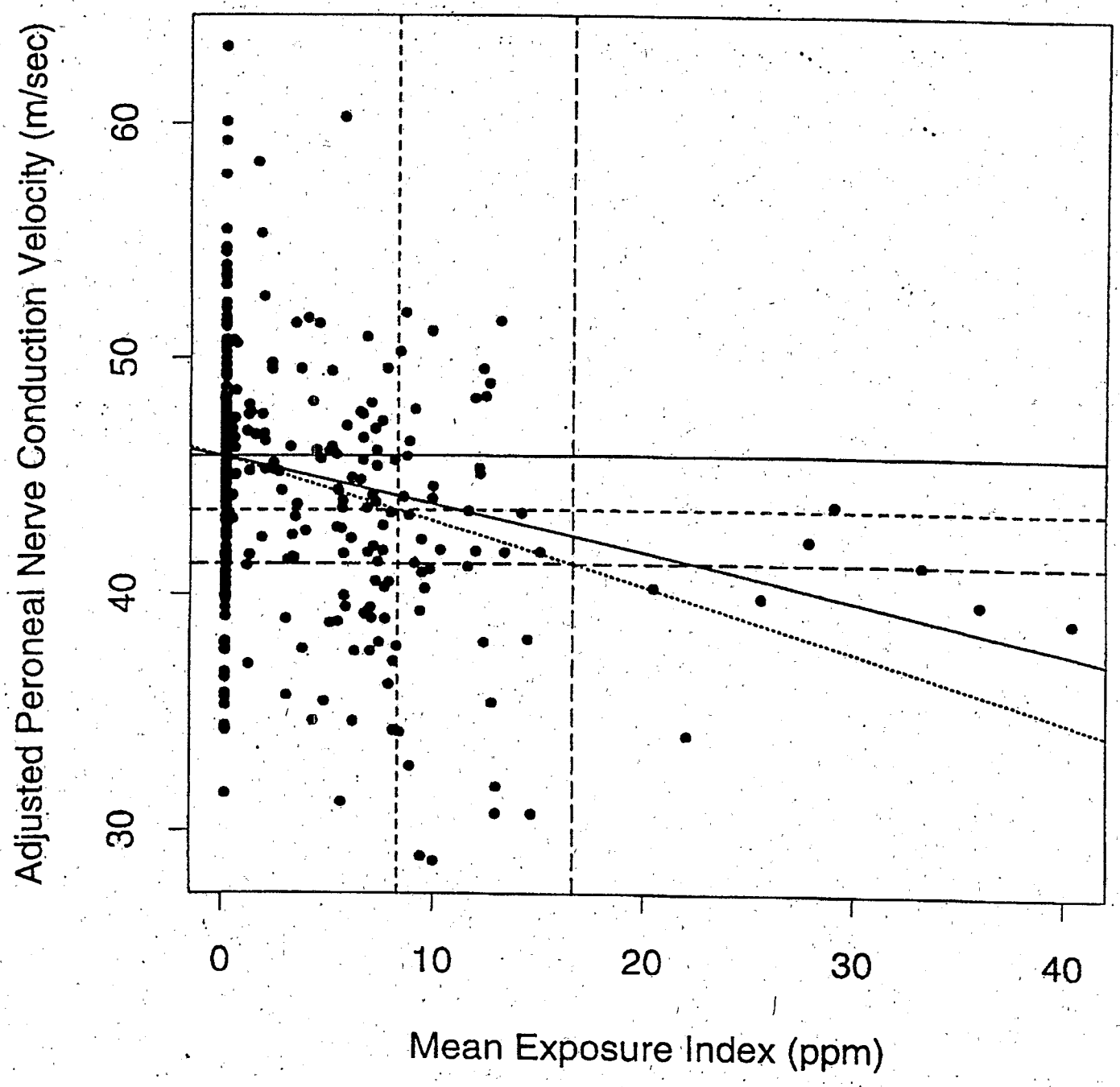


Figure 4. Carbon Disulfide Modeling (see text for explanation)

Example #2: 1,1,1,2-Tetrafluoroethane (HFC-134a)

Summary of Main Points Illustrated

Setting the BMR based on default and limit of detection approaches for dichotomous data.

Modeling a response with a high background incidence.

Comparison of Extra vs. Additional risk

Summary of Data

Study: Chronic rat inhalation
Endpoints modeled: Leydig cell hyperplasia, incidence
Endpoint used: same

The BMC for HCF-134a was based on the BMC calculated for testicular effects (Table 1) in a chronic study in rats (Hext and Parr-Dobrzanski, 1993; Collins et al., 1995).

Table 1. Data used for Benchmark Dose Analysis of HFC-134a

Concentration ppm	N	Incidence
0	85	27
2,500	79	25
10,000	85	31 (NOAEL)
50,000	85	40 (LOAEL)

Selection of BMR

Default BMR - The benchmark response for Leydig cell hyperplasia was defined as a 10% extra risk of response.

Selection of BMR based on Limit of Detection. - Use of the limit of detection approach significantly changes the BMC for this chemical because of the high background and the very shallow dose-response curve.

The determination of a "typical" design that includes evaluation of Leydig cell hyperplasia in rats, and control incidences is a first step in determining the BMR using the limit of detection approach. Leydig cell hyperplasia would be observed during routine histopathology of the testes

in a subchronic or chronic animal study. The number of animals typically used in such studies are 10-20 for a subchronic study and 50-100 for a chronic study. To determine the appropriate BMR, a "typical sample size could be determined to be, e.g., n = 25. Alternatively, the sample size from the study could be used, n = 85. There is a provision for the latter approach in the guidelines if the endpoint is not typically examined and there is no historical information on a typical design and background incidence level.

Secondly, the "typical" background rates must be determined. If an additional risk formulation is used, background is relatively unimportant, but the sample size makes a big difference. Again, the choice is between the use of a "typical" background rate for the endpoint and the use of the background in the study. Because a review of the incidence of Leydig cell hyperplasia has not been done to determine the typical incidence and the variability among studies, the background observed in the study (approximately 30%) was used in the determination of the BMR.

Finally, the required power level to be used must be determined. For the purpose of this example, power levels of 50% and 80% will be used. The BMRs for various sample sizes, backgrounds, and power levels are shown in Table 2.

Table 2. Benchmark response levels for various sample size, background, and power.

		Extra Risk		Additional Risk	
SS	Background	50% Power	80% Power	50% Power	80% Power
25	.20	0.33	0.47	0.26	0.37
25	.30	0.39	0.55	0.27	0.38
25	.40	0.46	0.63	0.28	0.38
85	.20	0.15	0.23	0.12	0.18
85	.30	0.19	0.28	0.13	0.20
85	.40	0.23	0.34	0.14	0.20

Calculation of the Benchmark Concentration

A. Use of Default BMR

Model: Polynomial and Weibull models (with and without threshold term) agreed. Used commercially available software.

The RfC for HCF-134a was based on the BMC calculated for testicular effects (Table 1) in a chronic study in rats (Hext and Parr-Dobrzanski, 1993; Collins et al., 1995). Because of the high background, a substantial difference in the estimates of benchmark concentration occurs for

extra risk vs. additional risk models (15,600 vs. 11000 ppm), based on the default use of a BMR of 0.1 (Table 3). An extra risk model was selected as most appropriate based on the conservative assumption of independence of mechanisms causing the background response and the treatment-related response. Dose-response models for dichotomous data included a polynomial (multistage) and a Weibull form and either including or not including a model parameter for a background-intercept (sometimes referred to as a threshold parameter). Using any of these four model forms, an excellent model fit was obtained and the BMC estimates were the same after rounding to two significant figures.

Table 3. Benchmark Concentrations for Leydig Cell Hyperplasia for HFC-134a based on the default BMR of 0.1. (Polynomial and Weibull models, extra risk, no threshold)

Endpoint/ Model	BMR %	MLE ppm	BMC ppm
Polynomial, no thresh, extra	10	19300	11000
Polynomial, no thresh, additional	10	28600	15600
Weibull, no thresh, extra	10	19300	11000

B. Use of Limit of Detection

The limit of detection approach used considerably higher BMR levels (.30 compared to the default of .10) because of the high control response level and consequently, a greater treatment effect required to be statistically significant. BMCs resulting for the BMRs based on the limit of detection for a background rate of 30%, 50% or 80% power and extra or additional risk are shown in Table 4.

Table 4. Benchmark Concentrations and MLE Estimates for Various Sample Size, Background, and Power. The polynomial model with no threshold term was applied.

SS	Background	50% Power		80% Power	
		MLE	BMC	MLE	BMC
			Extra Risk		
25	.30	95600	52500	154000	62400
85	.30	40800	22600	63500	35200
			Additional Risk		
25	.30	97500	51000	158000	62600
85	.30	40900	21600	67100	35300

The BMC is 5-6 times higher when the BMR is calculated based on the limit of detection approach for n=25 and 2-3 times higher for n=85, and for a sample size of 25, is higher than the highest exposure level in the study (see figure 5, based on extra risk). This difference in the BMCs would be translated directly into the derivation of the RfC or other health exposure limit because there would be no difference in the application of uncertainty factors for the two approaches to selecting the BMR. It is also noteworthy that the limit of detection approach to determining the BMR eliminates the difference between the BMCs for additional vs. extra risk because BMRs corresponding to extra versus additional risk are used for the appropriate models.

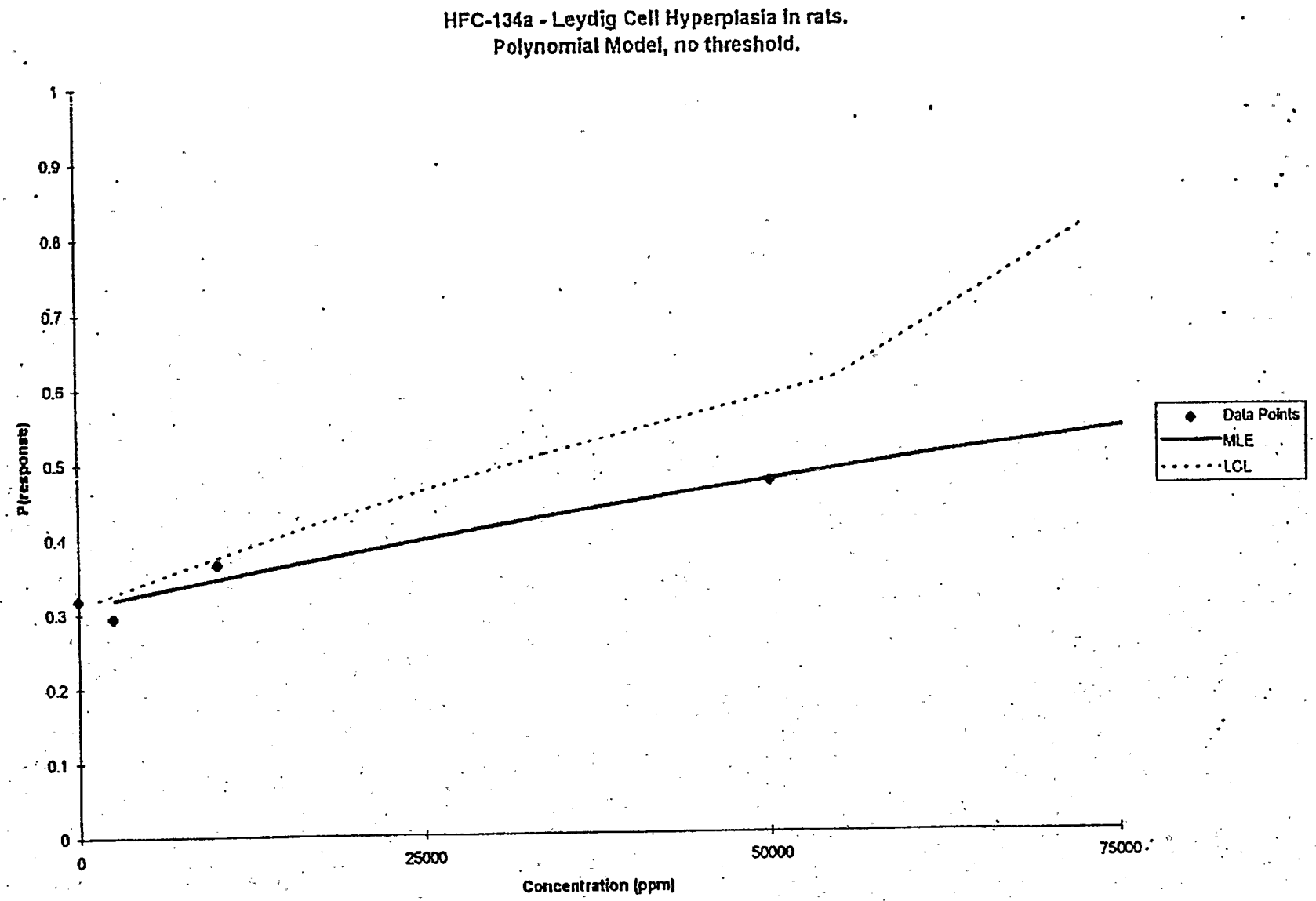


Figure 5. HFC-134 Modeling

Example #3: Boron

Summary of Main Points Illustrated

Combination of two experiments
Conversion of continuous data to dichotomous form.

Summary of Data Sets

Studies: Two prenatal developmental toxicity studies with boric acid administered in the diet (Heindel et al., 1992; RTI, 1994).
Endpoints modeled: fetal weight (as continuous and dichotomous), malformations, variations.
Endpoint used: fetal weight (continuous) because changes were seen at the lowest doses.

Table 1. Data used for Benchmark Dose Analysis for Boron - continuous data

Dose* (mg/kg/day)	Fetal Weight	SD	N (fetuses)
	Study A		
0	3.7	.32	218
78	3.45	.25	217
163	3.21	.26	205
330	2.34	.25	191
	Study B		
0	3.61	.24	211
19	3.56	.23	226
36	3.53	.28	220
55	3.50	.38	221
76	3.38	.26	236
143	3.16	.31	209

*The doses in the two studies are slightly different because they are based on food consumption measurements in the two studies.

Table 2. Data used for Benchmark Dose Analysis for Boron - dichotomized continuous data

Dose (mg/kg/day)	N	Incidence
	Study A	
0	431	21
78	432	51
163	408	152
330	386	384
	Study B	
0	416	21
19	460	24
36	437	42
55	437	40
76	471	70
143	411	162

Selection of the BMR

Benchmark response levels of 5% additional risk for dichotomous data, and control SD divided by 2 for continuous data were selected. The rationale for these choices is provided by Allen et al. (1996).

Calculation of the Benchmark Dose

Model: continuous power for continuous data, log-logistic (fetuses nested within litters) for dichotomous, additional risk.

The BMD for boron was derived by Allen et al. (1996). In this analysis, BMDs for several endpoints from two developmental studies in rats were calculated and compared. In addition, the analysis combined results from the 2 studies because they were done by the same lab, and because the 2nd study was done as a follow-up intended specifically to extend the dose range from the first study in order to clearly define a NOAEL. There are a few general issues illustrated by this analysis.

1. Combination of studies - Dose-response functions were fit to the individual data sets and these were compared using a likelihood ratio test. If this test indicated that the responses from the two studies were consistent with a single dose-response function, the model was fit to the combined data.

2. Fetal weights were analyzed in two ways. In the first, the average litter weights were modeled as a continuous variable (Figure 6). The BMR was defined as a 5% decrease in fetal weight from the control mean or as a decrease equal to the control standard deviation divided by 2. In the second approach, the individual litter weights were converted to dichotomous data using the 5th percentile in the corresponding control group as the definition of adversely affected pups. The model was then fit to the litter data expressed as probability of adverse response (Figure 7).

3. Combination of data on missing or malformed ribs - The analysis explored several approaches to weighting and combining data on alterations in thoracic rib XIII and in lumbar ribs.

Selection of the most appropriate BMD:

The fetal weight analysis using the continuous power model was recommended by Allen et al. (1996) for the following reasons:

- The BMD was lower than those calculated for rib effects or other malformations
- The analysis of fetal weight effects converted to dichotomous data showed that the dose-response pattern was not the same in the two studies so the studies could not be combined for this analysis.
- It is preferable to combine data from these two studies, if possible.

Table 3. Benchmark Dose Results for Boron

Analysis	MLE	BMD
Fetal weight - continuous data - BMR = 5% decrease		
Study A	80	56
Study B	68	47
Combined	78	59
Fetal weight - continuous data - BMR = decrease to control SD/2		
Study A	73	48
Study B	49	31
Combined	65	59
Fetal weight - dichotomized continuous data		
Study A	129	115
Study B	47	31
Combined	--	--

The BMDs from the combined continuous data using either a 5% decrease or a decrease to control SD/2 were the same, and were similar to the BMDs calculated for Studies A and B

alone, especially those based on a 5% decrease from the control mean. These results indicate that a BMD could be calculated without running the second study to define the NOAEL.

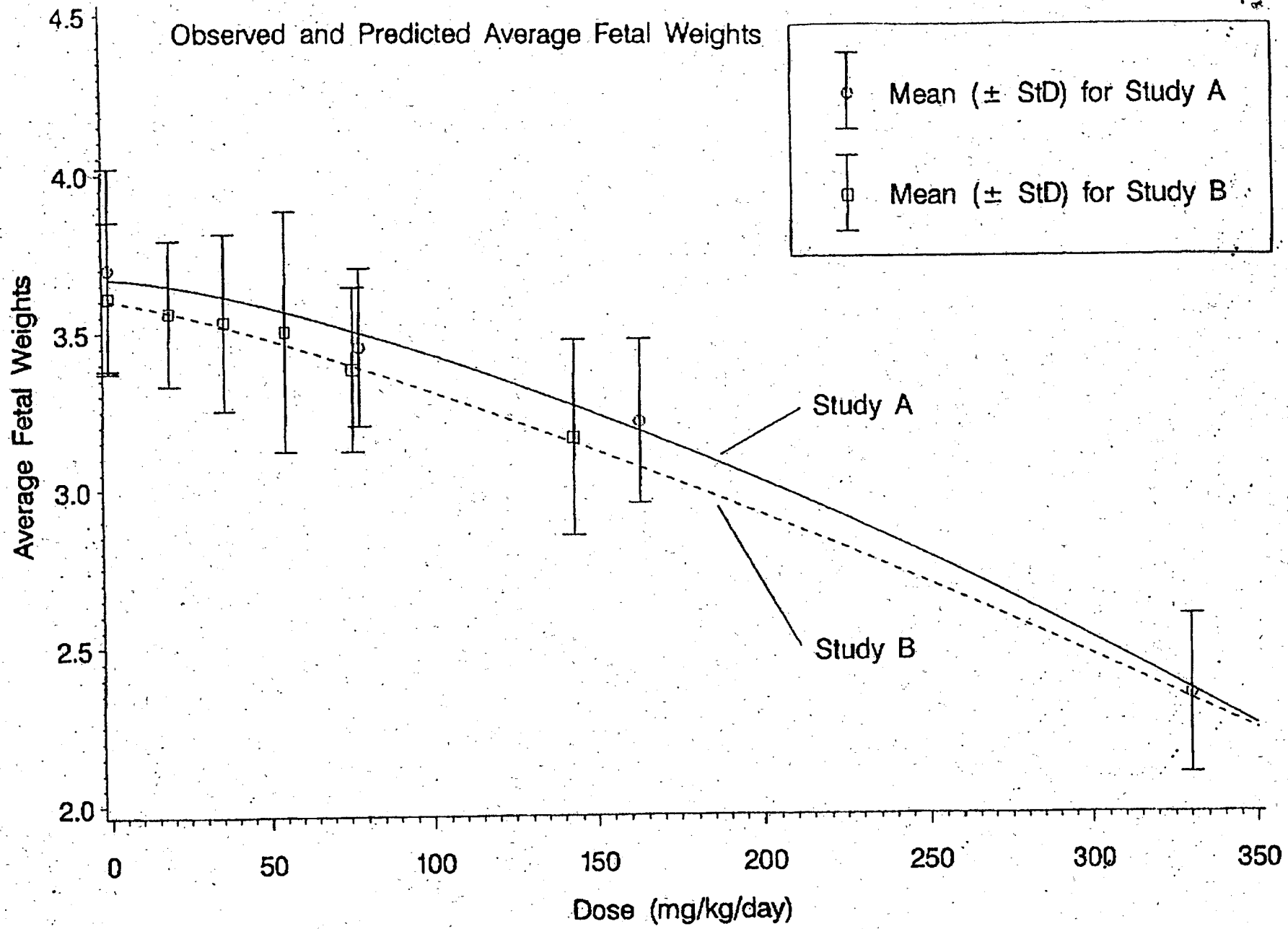


Figure 6. Boron Modeling - Continuous Data (from Allen et al., 1996)

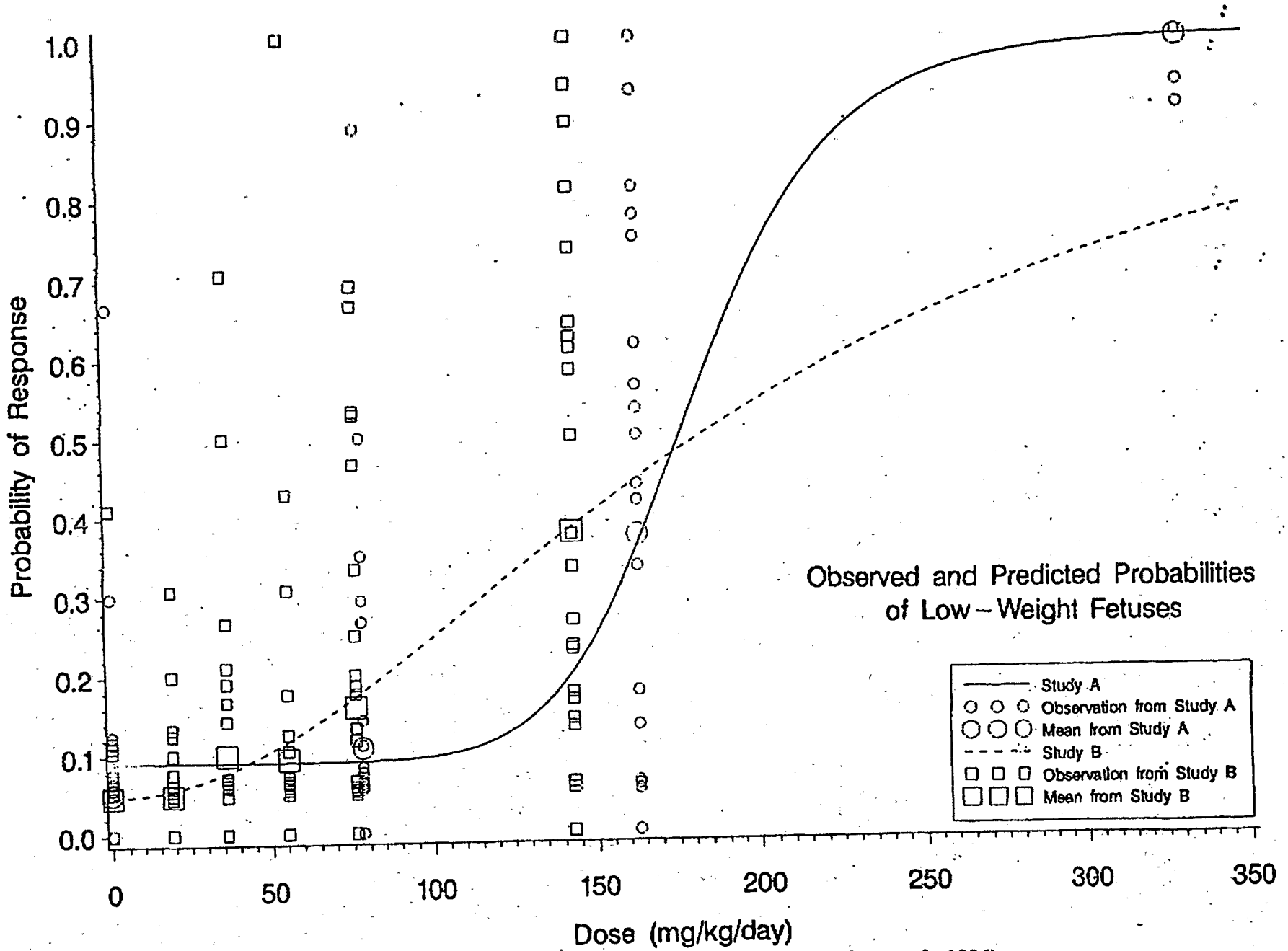


Figure 7. Boron Modeling - Dichotomous Data (from Allen et al., 1996)

Example #4: 1,3-Butadiene

Summary of Main Points Illustrated

BMC analysis with no NOAEL
What to do when model doesn't fit (drop the high dose level)

Summary of Data

Study: Chronic inhalation exposure study (NTP, 1991) in $B6C3F_1$ mice
Endpoints modeled: Ovarian, testicular and uterine atrophy
Endpoint used: Ovarian atrophy

Data from a chronic inhalation study were used to evaluate the noncancer health effects. Ovarian, testicular and uterine atrophy all showed effects related to exposure level, but ovarian atrophy was seen at all exposure levels, with no NOAEL. The data used are shown in Table 1.

Table 1. Data used in the benchmark concentration analysis for 1,3-butadiene.

	Ovarian Atrophy	
Exposure Level	No. Examined	% Affected
0.00	49	8.16
6.25 ppm	49	38.78
20.0 ppm	48	66.67
62.5 ppm	50	84.00
200.0 ppm	50	86.00
625.0 ppm	79	87.34

Selection of the BMR Level

The BMR was chosen as the default of 10% extra risk above controls (Kimmel et al. 1996). The limit of detection was determined based on the background rate (10%) and number of animals (50) in this study to compare with the default approach. Table 2 shows the BMR calculated on this basis for power levels of 50% and 80%.

Table 2. Determination of the BMR Using the Limit of Detection Approach

Background	N	Power	Extra Risk
.10	50	50%	.13
.10	50	80%	.22

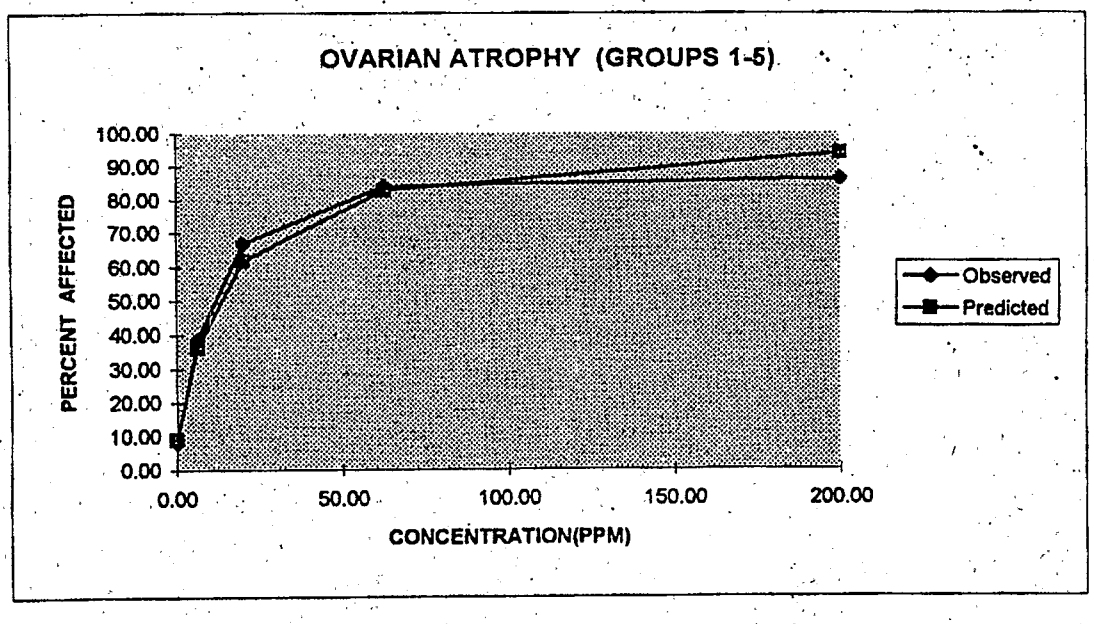
Calculation of the Benchmark Concentration

Model: log-logistic model used

Ovarian atrophy data could not be fit adequately using the quantal Weibull model. A log-logistic model was used to model the data.The model produced a poor fit for all 6 exposure groups, due to leveling off of the response at exposures above 62.5 ppm. Although the model was capable of fitting the first 5 exposure levels, the best fit of the data based on a graphical display of the fit was with exposure groups 1-4 (see Fig. 8). However, the BMCs based on the default of 10% risk were similar for exposure groups 1-5 and 1-4, and for additional risk and extra risk (Table 2). The BMCs based on the limit of detection for each power level also were similar for exposure groups 1-5 and 1-4, indicating that the model was giving a similar fit the data in the range of the BMR in both cases.

Table 2. BMCs for ovarian atrophy after 1,3-butadiene chronic inhalation exposure

LOAEL	Exposure Groups Modeled	BMC_{10}	BMC (based on limit of detection) 50%	80%
		Additional Risk		
6.25 ppm	1-5	1.23		
	1-4	1.03		
		Extra Risk		
	1-5	1.15	1.54	2.91
	1-4	0.96	1.29	2.43



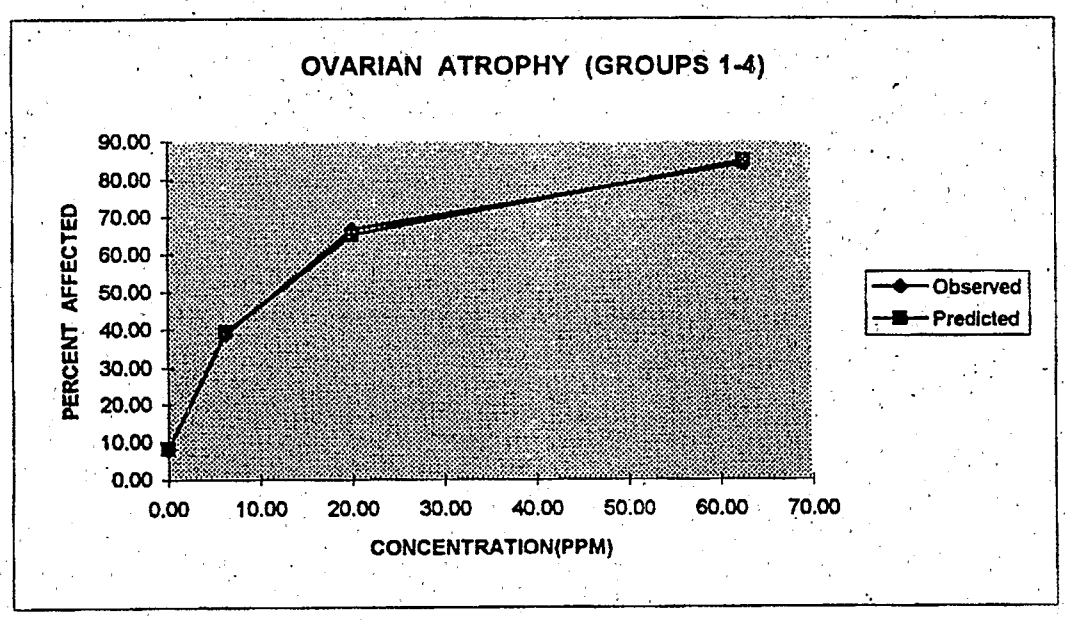


Figure 8. 1,3-Butadiene Modeling

APPENDIX E

LIST OF MODELS PLANNED TO BE INCLUDED IN FIRST RELEASE OF THE EPA SOFTWARE

BINARY REGRESSION MODELS

- Probit Model
- Weibull Model
- Logistic Model
- Gamma Multi-Hit Model
- Quantal Linear Model (One Hit)
- Quantal Quadratic Model
- Multistage Model (Quantal Polynomial)

NESTED BINARY REGRESSION MODELS

- Logistic Model
- Rai and Van Ryzin Model
- NCTR Model

CONTINUOUS REGRESSION MODELS

- Linear Model
- Polynomial Model
- Power Model