

1 **Environmental Health Criteria XXX**

2 **PRINCIPLES FOR MODELLING DOSE-RESPONSE FOR THE RISK**  
3 **ASSESSMENT OF CHEMICALS**

4  
5  
6  
7  
8  
9

10  
11 DRAFT

12  
13  
14  
15  
16  
17  
18

19 Prepared by the

20  
21

22 WHO TASK GROUP ON ENVIRONMENTAL HEALTH CRITERIA FOR PRINCIPLES FOR  
23 MODELLING DOSE-RESPONSE FOR THE RISK ASSESSMENT OF CHEMICALS

24  
25  
26 International Programme on Chemical Safety  
27 World Health Organization  
28 Geneva

29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41

42 February 2004

## TABLE OF CONTENTS

1	
2	
3	<b>FOREWORD</b>
4	
5	<b>LIST OF MEMBERS OF WHO PLANNING GROUP FOR IPCS HARMONIZATION</b>
6	<b>PROJECT ON DOSE-RESPONSE MODELLING</b>
7	
8	<b>LIST OF MEMBERS OF WHO TASK GROUP ON ENVIRONMENTAL HEALTH</b>
9	<b>CRITERIA FOR PRINCIPLES FOR MODELLING DOSE-RESPONSE FOR THE RISK</b>
10	<b>ASSESSMENT OF CHEMICALS</b>
11	
12	<b>LIST OF OTHER CONTRIBUTORS</b>
13	
14	<b>LIST OF SECRETARIAT</b>
15	
16	<b>LIST OF ACRONYMS AND ABBREVIATIONS</b>
17	
18	<b>1. INTRODUCTION, EXECUTIVE SUMMARY AND RECOMMENDATIONS</b>
19	1.1 Introduction
20	1.1.1 Intended Audience
21	1.1.2 Scope
22	1.2. Executive Summary
23	1.3 Conclusions and Recommendations
24	1.3.1 Conclusions
25	1.3.2 Recommendations
26	
27	<b>2. RISK ANALYSIS</b>
28	2.1 Decision Paradigms
29	2.2 Risk Analysis Paradigms
30	2.3 Motivations and Considerations for Producing a Formal Risk Assessment
31	2.3.1 Transparency and Justification
32	2.3.2 Public Health and Individual Health
33	2.3.3 Quantification and Computation
34	2.3.4 Cost of Assessment
35	2.4 Risk Assessment
36	2.4.1 Problem Formulation
37	2.4.1.1 Defining the Question
38	2.4.1.2 Prior Knowledge
39	2.4.1.3 Desired Outcomes
40	2.4.2 Risk Assessment Outcomes
41	
42	<b>3. DOSE-RESPONSE MODELLING - BASIC CONCEPTS</b>
43	3.1 Introduction
44	3.2 What Is Dose?
45	3.3 What Is Response?
46	3.4 What Is a Model?

1	3.5	What is Dose-Response Modelling?
2	3.6	Risk Versus Safety in Dose-Response Modelling
3	3.7	Summary
4		
5	<b>4.</b>	<b>DOSE-RESPONSE MODELLING - WHY AND WHEN TO USE</b>
6	4.1	Historical Perspectives
7	4.1.1	The NOAEL Approach to Acceptable/Tolerable Daily Intake
8		(ADI/TDI)
9	4.1.2	The Benchmark Dose Approach to Acceptable/Tolerable Daily
10		Intake (ADI/TDI)
11	4.2	Points of Consideration
12	4.2.1	General Aspects of Definition
13	4.2.2	Estimation Procedure
14	4.2.3	Uncertainty
15	4.2.4	Study Design
16	4.2.5	Biological Information
17	4.2.6	Comparison of Experimental Results
18	4.2.7	Risk Management Perspectives
19	4.3	Implementation Issues
20	4.4	Summary
21		
22	<b>5.</b>	<b>PRINCIPLES OF DOSE-RESPONSE MODELLING</b>
23	5.1	Data
24	5.1.1	Selection of Data
25	5.1.2	Data Types
26	5.2	Models and Distributions
27	5.2.1	Dose-Response Models
28	5.2.1.1	Continuous Dose-Response Models
29	5.2.1.2	Quantal Dose-Response Models
30	5.2.1.2	Thresholds
31	5.2.1.4	Modelling with Covariates
32	5.2.1.5	Biologically-Based Dose-Response Models
33	5.2.2	Statistical Distributions
34	5.2.2.1	Continuous Distributions
35	5.2.2.2	Discrete Distributions
36	5.3	Model Fitting and Estimation of Parameters
37	5.3.1	Criterion Function
38	5.3.2	Search Algorithms
39	5.4	Model Comparison
40	5.5	Representing Uncertainty
41	5.5.1	Sampling Error
42	5.5.2	Study Errors
43	5.5.3	Model Errors
44	5.6	Benchmark Dose and Benchmark Response Selection
45	5.7	Summary

1  
2 **6. COMMUNICATING RESULTS OF DOSE-RESPONSE MODELLING**

3 6.1 Introduction

4 6.1.1 Incorporation of the Outputs of DRM into Risk Assessments

5 6.1.2 Derivation of Health-Based Guidance Values

6 6.1.3 Estimation of the Margin of Exposure

7 6.1.4 Quantitative Estimations of the Magnitude of the Risk at Levels of  
8 Human Exposure

9 6.2 Presentation of Results

10 6.2.1 Tables

11 6.2.2 Graphs

12 6.3 Risk Assessment Context and Questions

13 6.4 Synopsis of Approach to Modeling

14 6.4.1 Data Sets

15 6.4.2 Severity

16 6.4.3 Uncertainty

17 6.5 Explaining/Interpreting the Output of the Dose-Response Analysis

18 6.5.1 Outputs in the Observable Biological Range

19 6.5.1.1 Health-Base Guidance Values

20 6.5.1.2 Margin of Exposure

21 6.5.2 Outputs Outside the Observable Biological Range

22 6.5.2.1 Prediction of Risks at Specified Exposure Levels

23 6.5.2.2 Prediction of Exposure Levels Producing Specified Risk  
24 Levels

25 6.5.2.3 Uncertainty Analyses

26 6.6 Issues for Risk Managers

27 6.6.1 Risk Assessment Issues

28 6.6.1.1 Population and Individual Effects

29 6.6.1.2 Risk Characterization

30 6.6.2 Risk Management Issues

31 6.6.2.1 Risk Management Options

32 6.6.2.2 Cost and Risk Benefit Analyses

33 6.6.2.3 Acceptable Level of Risk  
34

35 **REFERENCES CITED**

36  
37 **TERMINOLOGY**

## 1 **FOREWORD**

### 2 **Environmental Health Criteria**

#### 3 **Objectives**

4 In 1973 the WHO Environmental Health Criteria Programme was initiated with the following  
5 objectives:

- (i) to assess information on the relationship between exposure to environmental pollutants and human health, and to provide guidelines for setting exposure limits;
- (ii) to identify new or potential pollutants;
- (iii) to identify gaps in knowledge concerning the health effects of pollutants;
- (iv) to promote the harmonization of toxicological and epidemiological methods in order to have internationally comparable results.

6 The first Environmental Health Criteria (EHC) monograph, on mercury, was published in 1976  
7 and since that time an ever-increasing number of assessments of chemicals and of physical  
8 effects have been produced. In addition, many EHC monographs have been devoted to  
9 evaluating toxicological methodology, e.g. for genetic, neurotoxic, teratogenic and nephrotoxic  
10 effects. Other publications have been concerned with epidemiological guidelines, evaluation of  
11 short-term tests for carcinogens, biomarkers, effects on the elderly and so forth.

12 Since its inauguration the EHC Programme has widened its scope, and the importance of  
13 environmental effects, in addition to health effects, has been increasingly emphasized in the total  
14 evaluation of chemicals.

15 The original impetus for the Programme came from World Health Assembly resolutions and the  
16 recommendations of the 1972 UN Conference on the Human Environment. Subsequently the  
17 work became an integral part of the International Programme on Chemical Safety (IPCS), a  
18 cooperative programme of UNEP, ILO and WHO. In this manner, with the strong support of the  
19 new partners, the importance of occupational health and environmental effects was fully  
20 recognized. The EHC monographs have become widely established, used and recognized  
21 throughout the world.

22 The recommendations of the 1992 UN Conference on Environment and Development and the  
23 subsequent establishment of the Intergovernmental Forum on Chemical Safety with the priorities  
24 for action in the six programme areas of Chapter 19, Agenda 21, all lend further weight to the  
25 need for EHC assessments of the risks of chemicals.

26 This EHC on Principles for Modelling Dose-Response for the Risk Assessment of Chemicals is  
27 part of the IPCS project on the Harmonization of Approaches to the Assessment of Risk from

1 Exposure to Chemicals (Harmonization Project). The Intergovernmental Forum on Chemical  
2 Safety (IFCS) Forum III, held in Salvador da Bahia in October 2000, agreed on Priorities for  
3 Action Beyond 2000, which further defined the actions recommended to be taken. Forum III  
4 declared that by 2004, IPCS and the Inter-Organization Programme for the Sound Management  
5 of Chemicals (IOMC, which comprises seven intergovernmental organizations) should have  
6 ensured that recommendations for harmonized assessment approaches were available for  
7 terminology, cancer, and reproductive and developmental toxicology and that common principles  
8 for the assessment approach to other specific toxicological end-points, such as  
9 immunotoxicology, endocrine disruptors, and ecotoxicology, should be adopted wherever  
10 possible.

11 The IPCS Harmonization Project, which is ongoing, states that “harmonization,” in the context  
12 of chemical risk assessment, should not simply be equated with standardization. It is not a goal  
13 of the project to standardize risk assessments globally, as that is considered to be neither  
14 appropriate nor feasible. Instead, harmonization is thought of as an effort to strive for  
15 consistency among approaches and to enhance understanding of the various approaches to  
16 chemical risk worldwide. Thus, harmonization is defined, in a step-wise fashion, as an  
17 understanding of the methods and practices used by various countries and organizations so as to  
18 develop confidence in, and acceptance of, assessments that use different approaches. It further  
19 involves a willingness to work towards convergence of these approaches or methods as a longer-  
20 term goal.

21 Achieving harmonization of approaches is considered to provide a framework for comparing  
22 information on risk assessment; understanding of the basis for exposure standards for specific  
23 chemicals in different countries; savings of time and expense by sharing information and  
24 avoiding duplication of work; and credible science through better communication among  
25 organizations and peer review of assessments and assessment procedures. The stated  
26 Harmonization Project mission is to ensure better chemical risk assessment and hence  
27 management practices that promote the protection of human health and the environment within  
28 the framework of sustainable development.

29 This ongoing project is overseen by a geographically representative Harmonization Project  
30 Steering Committee and a number of ad hoc Working Groups that manage the detailed work.  
31 Finalization of documents includes a rigorous process of international peer review and public  
32 comment.

### 33 **Scope**

34 The criteria monographs are intended to provide critical reviews on the effect on human health  
35 and the environment of chemicals and of combinations of chemicals and physical and biological  
36 agents. As such, they include and review studies that are of direct relevance for the evaluation.  
37 However, they do not describe *every* study carried out. Worldwide data are used and are quoted  
38 from original studies, not from abstracts or reviews. Both published and unpublished reports are  
39 considered and it is incumbent on the authors to assess all the articles cited in the references.  
40 Preference is always given to published data. Unpublished data are used only when relevant  
41 published data are absent or when they are pivotal to the risk assessment. A detailed policy

1 statement is available that describes the procedures used for unpublished proprietary data so that  
2 this information can be used in the evaluation without compromising its confidential nature  
3 (WHO (1999) Guidelines for the Preparation of Environmental Health Criteria. PCS/99.9,  
4 Geneva, World Health Organization).

5 In the evaluation of human health risks, sound human data, whenever available, are preferred to  
6 animal data. Animal and *in vitro* studies provide support and are used mainly to supply evidence  
7 missing from human studies. It is mandatory that research on human subjects is conducted in full  
8 accord with ethical principles, including the provisions of the Helsinki Declaration.

9 The EHC monographs are intended to assist national and international authorities in making risk  
10 assessments and subsequent risk management decisions. They represent a thorough evaluation of  
11 risks and are not, in any sense, recommendations for regulation or standard setting. These latter  
12 are the exclusive purview of national and regional governments.

### 13 **Procedures**

14 The Dose-Response Modelling Planning Group met on 10 October 2002 in Geneva to develop an  
15 outline and proposed timeframe for the project. A first draft working paper, including  
16 contributions from several additional authors, was prepared by Drs C. Carrington and M. Bolger  
17 and distributed to the Task Group prior to the Task Group Meeting from 13 - 17 September 2004.  
18 The Task Group members serve as individual scientists, not as representatives of any  
19 organization, government or industry. All individuals who as authors, consultants or advisers  
20 participate in the preparation of the EHC monograph must, in addition to serving in their  
21 personal capacity as scientists, inform the RO if at any time a conflict of interest, whether actual  
22 or potential, could be perceived in their work. They are required to sign a conflict of interest  
23 statement. The Chairpersons of Task Groups are briefed before each meeting on their role and  
24 responsibility in ensuring that these rules are followed. Such a procedure ensures the  
25 transparency and probity of the process. Their function is to evaluate the accuracy, significance  
26 and relevance of the information in the document. A summary and recommendations for further  
27 research and improved safety aspects are also required. The composition of the Task Group is  
28 dictated by the range of expertise required for the subject of the meeting and, where possible, by  
29 the need for a balanced geographical distribution. The first draft working paper was revised  
30 during the Task Group Meeting and during a subsequent internal Task Group internet forum.

31 *(This draft will be available on the IPCS website for external review and comment for a period*  
32 *of two months. These comments will be reviewed by a Task Group and, as necessary, additions*  
33 *and revisions to the second draft working paper will be made. A comprehensive file of all*  
34 *comments received on the draft will be maintained and made available on request. When this*  
35 *Task Group has completed its review and the WHO Secretariat is satisfied as to the scientific*  
36 *correctness and completeness of the document, it will then go for language editing, reference*  
37 *checking and preparation of camera-ready copy. At this time a copy of the final draft is sent to*  
38 *the Chairperson and Rapporteurs of the Task Group to check for any errors. After approval by*  
39 *the Director, IPCS, the monograph will be submitted to the WHO Office of Publications for*  
40 *printing and will be available on the IPCS website.)*

- 1 The financial support for this project by Health Canada, the Ministry of Health of Japan, the UK
- 2 Food Standards Agency, and the US National Institute of Environmental Health Sciences in
- 3 gratefully acknowledged.

1 WHO PLANNING GROUP FOR IPCS HARMONIZATION PROJECT ON DOSE-  
2 RESPONSE MODELLING

3 ***Members***

4 Dr P. Bolger, US Food and Drug Administration, College Park, MD, USA

5

6 Prof E. Dybing, Norwegian Institute of Public Health, Oslo, Norway

7

8 Dr L. Edler, German Cancer Research Center, Heidelberg, Germany

9

10 Dr M. Hartley, National Industrial Chemicals Notification and Assessment Scheme (NICNAS),  
11 Sidney, Australia

12

13 Dr A. Knaap, National Institute of Public Health and the Environment (RIVM), Bilthoven,  
14 The Netherlands

15

16 Dr C. Portier, US National Institute for Environmental Health Sciences (NIEHS), Research  
17 Triangle Park, NC, USA

18

19 Prof A. Renwick, University of Southampton, Southampton, United Kingdom

20

21 ***Secretariat***

22 Dr T. Damstra, International Programme on Chemical Safety, World Health Organization,  
23 Research Triangle Park, NC, USA

24 Dr S Page, International Programme on Chemical Safety, World Health Organization, Geneva,  
25 Switzerland

26 Ms C. Sonich-Mullin, International Programme on Chemical Safety, World Health Organization,  
27 Cincinnati, OH, USA

28 Ms C. Vickers, International Programme on Chemical Safety, World Health Organization,  
29 Geneva, Switzerland

30

31

32

1  
2 WHO TASK GROUP ON ENVIRONMENTAL HEALTH CRITERIA FOR PRINCIPLES FOR  
3 MODELLING DOSE-RESPONSE FOR THE RISK ASSESSMENT OF CHEMICALS

4 *List of Members*

- 5 Dr P. Bolger, US Food and Drug Administration, College Park, MD, USA  
6  
7 Prof A. Boobis, IMPERIAL College London, United Kingdom  
8  
9 Dr C. Carrington, US Food and Drug Administration, College Park, MD, USA  
10  
11 Dr V. Cogliano, International Agency for Research on Cancer (IARC), Lyon, France  
12  
13 Prof E. Dybing, Norwegian Institute of Public Health, Oslo, Norway  
14  
15 Dr L. Edler, German Cancer Research Center, Heidelberg, Germany  
16  
17 Prof E. Faustman, University of Washington, Seattle, WA, USA  
18  
19 Mr J. Howlett, Food and Agriculture Organization of the United Nations, Rome, Italy  
20  
21 Dr M. Healy, Food Standards Australia New Zealand, Canberra, BC, Australia  
22  
23 Dr J. Kleiner, European Food Safety Authority, Brussels, Belgium  
24  
25 Dr A. Knaap, National Institute of Public Health and the Environment (RIVM), Bilthoven,  
26 The Netherlands  
27  
28 Dr J. Larsen, Danish Institute of Food and Veterinary Research, Soborg, Denmark  
29  
30 Dr D. Lovell, University of Surrey, Surrey, United Kingdom  
31  
32 Dr C. Portier, US National Institute for Environmental Health Sciences (NIEHS), Research  
33 Triangle Park, NC, USA  
34  
35 Prof A. Renwick, University of Southampton, Southampton, United Kingdom  
36  
37 Prof T. Sanner, Institute for Cancer Research, Oslo, Norway  
38  
39 Dr J. Schlatter, Swiss Federal Office of Public Health, Zürich, Switzerland  
40  
41 Dr R. Setzer, Jr., US Environmental Protection Agency, Research Triangle Park, NC, USA  
42  
43 Prof W. Slob, Institute of Public Health and the Environment (RIVM), Bilthoven,

1 The Netherlands

2

3 Dr A. Wadge, Food Standards Agency, London, United Kingdom

4

5 Prof G. Williams, New York Medical College, Valhalla, NY, USA

6

7

8 ***List of Other Contributors***

9

10 Prof E Calabrese, University of Massachusetts, Amherst, MA USA

11

12

13 ***Secretariat***

14 Dr S Page, International Programme on Chemical Safety, World Health Organization, Geneva,

15 Switzerland

16 Dr A Tritscher, International Programme on Chemical Safety, World Health Organization,

17 Geneva, Switzerland

18

19

1

2

**ACRONYMS AND ABBREVIATIONS**

3

4

5 ADI acceptable-daily-intake

6

7 BMD benchmark dose

8

9 BMDL benchmark dose lower confidence limit

10

11 BMR benchmark response

12

13 CSAF chemical-specific adjustment factor

14

15 DR dose-response

16

17 DRM dose-response modelling

18

19 EC European Commission

20

21 ECH Environmental Health Criteria

22

23 FAO Food and Agriculture Organization of the United Nations

24

25 IARC International Agency for Research on Cancer

26

27 IPCS International Programme on Chemical Safety

28

29 JECFA Joint FAO/WHO Expert Committee on Food Additives

30

31 LOAEL lowest-observed-adverse-effect-level

32

33 MOE margin-of-exposure

34

35 NOAEL no-observed-adverse-effect-level

36

37 NAS National Academy of Sciences (USA)

38

39 NTP National Toxicology Program (USA)

40

41 OECD Organization for Economic Co-operation and Development

42

43 OSHA Occupational Safety and Health Administration (USA)

44

45 TDI tolerable-daily-intake

1		
2	UF	Uncertainty Factor
3		
4	USEPA	United States Environmental Protection Agency
5		
6	WTO	World Trade Organization
7		
8	WHO	World Health Organization

## 1. INTRODUCTION, EXECUTIVE SUMMARY AND RECOMMENDATIONS

### 1.1 Introduction

The International Programme on Chemical Safety (IPCS) and other public health organizations have recognized the importance of the harmonization of procedures to enhance the quality of risk assessments, to improve the transparency of the risk assessment process and to facilitate risk communication.

Public health decisions on the plausible risks of chemical exposures can include several possible outcomes. The ultimate goal is to implement a risk management action that will produce the desired reduction of risk. Among the first objectives of a risk assessment is the determination of the presence or absence of a cause-effect relationship. If there is sufficient plausibility for the presence of such a relationship, then dose-response (DR) information is needed.

Extrapolation is a fundamental problem in the quantitative health risk assessment of exposures to chemicals that produce toxicity in experimental systems. Adverse health effects of chemicals are typically evaluated in laboratory animals at significantly higher doses than the levels at which humans may be exposed. Also, for certain substances for which the exposure can be controlled, such as food additives and residues of pesticides and veterinary drugs, the quantitation of the risk above the level of exposure that has been assessed to be safe (such as the acceptable daily intake (ADI)) can be difficult. This is particularly true in cases of temporary excursions above an ADI.

The use of mathematical and statistical approaches in hazard characterization is increasing. While DR models have been available for some time, their use has been somewhat limited either because of a lack of appropriate scientific information or agreed upon approaches and methods for how to obtain and use available DR information appropriately. Dose-response modelling (DRM) involves a number of choices based upon scientific experience, data availability and mathematical tractability, and can take on many different forms and be used in many different ways. A recent review of the available quantitative approaches for hazard characterization noted that mathematical modelling of the dose-response relationship could improve the risk assessment process (Edler et al., 2002).

DR models may improve and generate more reliable predictions, but they can never be proved to be completely correct. Therefore, it is necessary to rely on scientific judgement to determine the utility of risk predictions from DRM in making public health decisions. It is important to remember that risk numbers derived from DRM can be misleading for a variety of reasons and, like any other tool used in science, needs to be utilized in a broader context of all of the available scientific knowledge. The final standard that prevails in risk assessment is biological plausibility, rather than mathematical correctness or statistical certainty. It is this inherent uncertainty and the communication thereof for which modelling and quantitative risk assessment can be particularly valuable.

### 1.1.1 Background

This Environmental Health Criteria report (EHC) is intended primarily to provide descriptive guidance for risk assessors in using DRM in hazard characterization. It will also provide mathematical modellers an appreciation of issues of modelling in the context of the risk assessment process. Risk managers will be able to obtain a general understanding of the applications and limitations of DRM. For both risk assessors and risk managers, some considerations for communicating the results of risk assessments that use DRM are presented.

### 1.1.2 Scope

This EHC is part of the on-going review of the underlying scientific bases for decision-making in chemical risk assessment by IPCS. It involves specific consideration of the area of DR assessment in the evaluation of information from toxicological studies in animals and human clinical and epidemiological studies and does not include consideration of other aspects of quantitative risk assessment, such as physiologically-based modelling.

The discussions are concerned with that subset of cause-effect relationships commonly referred to as DR models, that are typically used to characterize the biological effects of intentional (e.g., drugs and nutrients) and unintentional (e.g., contaminants) exposure to chemicals. DR models are commonly used in microbiological risk assessments (e.g., WHO, 2004a).

This EHC is intended to provide guidance in a number of areas relevant to DRM. Initially there is a discussion of the risk analysis paradigm (Chapter 2) and the basic concepts of DRM (Chapter 3). In Chapter 4, the use of DRM is described, including comparing the No Observed Adverse Effect Level (NOAEL) with the Bench Mark Dose (BMD) modelling method. Chapter 5 provides the principles for DRM, including data considerations, model descriptions, model fitting and parameter estimation, model comparisons, and uncertainty. This chapter also includes discussion of BMD approaches. Chapter 6 discusses the provision of the scientific advice by risk assessors to the managers. This final chapter includes an explanation of the output of the dose-response analysis and the strengths and weaknesses of DRM.

There is only limited treatment of the mathematical and statistical considerations for DRM. References and links are provided for more in-depth treatments, modelling tools, and examples.

## 1.2 Executive Summary

DRM, for use in quantitative risk assessment and ultimately for informing public health decisions about chemical exposures, can be described as a six-step process. The first four steps of data selection, model selection, statistical linkage, and parameter estimation constitute dose-response analysis. These steps relate to the process through which a mathematical description of the data is obtained in order to evaluate predicted responses for known doses or to obtain dose estimates when a chosen response is of interest. The fifth step involves the integration of the results of the dose-response analysis with estimates of exposure for the purposes of guiding public health decisions. The final step, which can optionally be applied earlier in DRM, involves

1 an assessment of the quality of the does-response analysis and the sensitivity of model  
2 predictions to the assumptions used in the analysis.

3  
4 The characterization of DR relationships in animal and human studies has been a major  
5 component of hazard characterization and has been used in the extrapolation of incidences of  
6 adverse effects in the range of human exposure levels. Over the years a variety of methods have  
7 been developed to accommodate such relationships, including the no-observed-adverse-effect-  
8 level (NOAEL) and acceptable daily intake (ADI), as well as other extrapolation methods for  
9 improving extrapolation to low doses. DRM may prove useful in risk assessments when a  
10 sufficient amount of animal or human dose response data is available and the uncertainties  
11 associated with such an extrapolation are deemed to be tenable. The standard NOAEL approach,  
12 a special case of DRM, identifies a single dose which is assumed to be without an appreciable  
13 adverse effect, while DRM estimates in general are based on data from the entire does-response  
14 curve for the critical effect. The DR model reflects the characteristics of the dose-response curve,  
15 particularly in providing estimates of the slope, and in the case of a regression framework, it  
16 provides standard error and confidence intervals for the model parameters. A disadvantage of  
17 using the NOAEL approach is that it is not possible to quantify the degree of variability and  
18 uncertainty that may be present, while other DR models can facilitate the analysis of sensitivity  
19 and uncertainty. Consideration of a DR model can optimize study design and clarify the need for  
20 additional studies. The NOAEL approach incorporates biological information through the  
21 application of 'expert' but subjective judgement. DRM has the potential for a more 'science rich'  
22 analysis through the more formal quantitative inclusion (e.g., factors, covariates) into the models.  
23 Estimates derived from DRM enhance the ability to quantitatively compare different experiments,  
24 effects and compounds within a common framework. DRM can enhance risk and safety  
25 assessments as well as provide opportunities to consider the likelihood of effects outside the  
26 observable range.

27  
28 The choice of the models to be used depends upon the type of data. They should include a model  
29 for DR and a model for the variability of the data. Once models are fit to a data set, the degree to  
30 which they individually describe the data can be evaluated using goodness of fit measures. In  
31 addition, their ability to describe the data with respect to each other may be compared.  
32 Uncertainties about the inferences that result from such models fall into three main categories:  
33 statistical uncertainty of inferences due to variability among the responses of experimental  
34 subjects, variability among experiments due to unavoidable differences in experimental  
35 execution and uncertainty due to the fact that the true model for the data is unknown. DR  
36 analysis needs to address all three sources of variability and uncertainty whenever possible. One  
37 particularly important application of DRM is the calculation of benchmark doses (BMD). These  
38 are doses at which it is inferred that a particular level of response would occur. When data are  
39 available, BMDs are an alternative to the NOAEL approach in the calculation of health guidance  
40 values, such as ADIs or Tolerable Daily Intakes (TDIs). When extrapolation is necessary, the  
41 uncertainty associated with a prediction should be represented. It is often especially important to  
42 include model uncertainty.

43  
44 DRM offers the potential to provide additional information for the risk manager. The output of  
45 DRM should be directed towards addressing specific questions about the likelihood of adverse  
46 health effects. It can be presented in three principal ways. Firstly, it can be used for the

1 establishment of a health-based guidance value, such as an ADI or TDI, analogous to current  
 2 procedures based on a NOAEL or lowest-observed-adverse-effect-level (LOAEL). DRM can be  
 3 a more scientifically robust method for determining health-based guidance values. Secondly, the  
 4 output from DRM can be used in risk management to estimate a margin of exposure (MOE) by  
 5 calculation of the ratio of the dose corresponding to a given limit of response and a human  
 6 exposure level. Thirdly, based on the modelled DR relationship, the output can be a quantitative  
 7 estimation of the magnitude of the risk at the level of human exposure. DRM can provide better  
 8 information on the likelihood of effects at low doses that are below the levels observed in  
 9 biological systems and can also provide better estimates of the statistical uncertainties  
 10 surrounding estimates of likely effects. There are several factors that can impact the type of  
 11 outputs from DRM exercises and may be of importance to the risk manager. These include  
 12 multiple data sets, information on severity, and uncertainties. DRM can be used with exposure  
 13 data to identify sub-populations at risk. DRM can also be used to assist risk managers in  
 14 determining priorities and evaluating the consequences of proposed interventions aimed at  
 15 reducing the risk. For risk communication, the use of DRM and other probabilistic assessment  
 16 techniques offers opportunities and challenges. DRM evaluations can produce information in  
 17 several formats including DR functions that allow, along with estimates of exposure, for the  
 18 prediction of risks at specified exposure levels and functions allowing for the estimation of  
 19 exposure levels resulting in specified risks. This includes estimates of the possible risk at intakes  
 20 above a health-based guideline, use as an ADI. They also offer approaches to compare  
 21 competing risk or benefits and provide a focus on uncertainties that can influence the predicted  
 22 risk. However, unless the situation of risk is viewed at the population level, there is the  
 23 conundrum that while one may be able to explain the level of risk in those circumstances where  
 24 there is no safe level of exposure, this can result in informing some percentage of the population  
 25 that they are predicted to experience some effect deemed to be adverse. It must be recognized  
 26 that the use of DRM requires a certain quantity and quality of data, as well as specific expertise.  
 27 Thus, DRM may not always be possible.

28  
 29 The potential 'on-going' use of the estimates from DRM can, from a risk management  
 30 perspective, give an improved characterization for decision making by

- 31
- 32 • providing information about what happens above the safety level (magnitude and types  
 33 of health impacts);
  - 34
  - 35 • showing benefits from different regulatory actions;
  - 36
  - 37 • giving the decision maker a 'more-than-one-point' appreciation of the data;
  - 38
  - 39 • promoting consistency in decisions, if appropriate adjustments are made for differences  
 40 in effect, effect level, species, and study design; and
  - 41
  - 42 • allowing for an iterative interaction between the risk assessor and manager on a  
 43 continuous and ongoing basis.
  - 44

45 The use of DRM and probabilistic assessment techniques to quantitatively describe variability  
 46 and uncertainty brings new challenges in risk communication. Some of these challenges are

- 1
- 2 • explaining that some percentage of the population is predicted to exceed the safety level
- 3 and/or experience an adverse effect;
- 4
- 5 • explaining the level of risk in those circumstances where there is assumed to be no safe
- 6 level of exposure;
- 7
- 8 • comparing competing risks or benefits;
- 9
- 10 • providing a focus on uncertainties that influence the predicted risk; and
- 11
- 12 • explaining that a risk estimate pertains to what may occur in a population level and not
- 13 that which pertains to the individual, and noting that this is also the case for the
- 14 ADI/TDI approach.
- 15

### 16 **1.3 Conclusions and Recommendations**

#### 17 **1.3.1 Conclusions**

- 18
- 19
- 20 • The use of DRM becomes relevant for hazard characterization when there is a
- 21 need to base the starting point for the low dose extrapolation on more dose-
- 22 response information than on that represented by a single dose-response point
- 23 such as the LOAEL, and in particular, when a quantification of the dose-response
- 24 relationship of the observed data is required. DRM can be considered when a
- 25 sufficient amount of dose response information is available, either from an
- 26 experimental animal bioassay or from a human study.
- 27
- 28 • DRM may not be necessary or appropriate for all risk assessments. Where the
- 29 exposure can be controlled and adequate information for the risk assessment is
- 30 generally provided prior to marketing, (e.g., for food additives, pesticides and
- 31 veterinary drugs,) the derivation of an ADI is almost always the appropriate
- 32 approach..
- 33
- 34 • Application of DRM for all endpoints can be cost-prohibitive, so the data sets
- 35 selected should generally reflect the more sensitive endpoints available. However,
- 36 it is sometimes very difficult to identify the most sensitive endpoints without
- 37 modelling all of them.
- 38
- 39 • The use of BMD approaches is preferable in the observed dose range if there are
- 40 sufficient data for their application.
- 41
- 42 • The Margin of Exposure (MOE) approach has been used in risk management by
- 43 some institutions and countries to provide an estimate of the level of exposure
- 44 (dose) of compounds that are genotoxic and carcinogenic that would result in an
- 45 acceptable level of risk.
- 46

- 1           • Outside the observable biological range, outputs of DRM evaluations can produce  
2 risk information in several formats including dose response functions that allow  
3 for the prediction of risks at specified exposure levels and functions allowing for  
4 the estimation of exposure levels resulting in specified risks. In addition, DRM  
5 exercises can provide uncertainty analyzes and the identification of the major  
6 factors contributing to uncertainty in the estimates of risk.  
7

### 8           **1.3.2 Recommendations**

9

- 10           • Better guidance for combined analysis of different populations for more precisely  
11 estimating BMDs needs to be developed.  
12
- 13           • Toxicity testing protocols [e.g., Organization for Economic Co-operation and  
14 Development (OECD) guidelines] should be reviewed for optimization for BMD  
15 and other DRM approaches, including optimal designs for the number animals  
16 and number of doses for different DR curves. Additional research is needed for  
17 the development of optimal study designs.  
18
- 19           • Additional research is needed to determine the biological basis for extrapolation,  
20 e.g., biomarkers, tumour precursors, genetically modified animals, and  
21 toxicokinetics for target dose estimation.  
22
- 23           • Better understanding of when and how to use the Benchmark Response Level  
24 (BRL) needs to be developed.  
25
- 26           • Better understanding of the shape of the DR curve at low doses needs to be  
27 developed.  
28
- 29           • Improved guidance for risk communication based on the results of DRM and  
30 other probabilistic assessment techniques needs to be developed. This should  
31 include communication of the types of uncertainty and the relation to statistical  
32 variability, imprecision, and the use of confidence intervals.  
33
- 34           • The use of DRM should be reviewed and additional general principles for its use  
35 developed when additional experience becomes available. Links to case studies  
36 of risk assessments using DRM are (*will be*) available at  
37 <http://www.who.int/ipcs/methods/en/>.

## 2. RISK ANALYSIS

### 2.1 Decision Paradigms

A risk analysis decision paradigm is a formal representation of a process which distinguishes the scientific bases from the risk management objectives and generally contains a component where the probability of harm is estimated. This component of the decision paradigm is referred to as the risk assessment. As a probability calculation, a risk assessment will include both a statement of the objective under consideration (i.e., the harm) and the basis for the assertion that the harm may occur (i.e., the probability).

### 2.2 Risk Analysis Paradigms

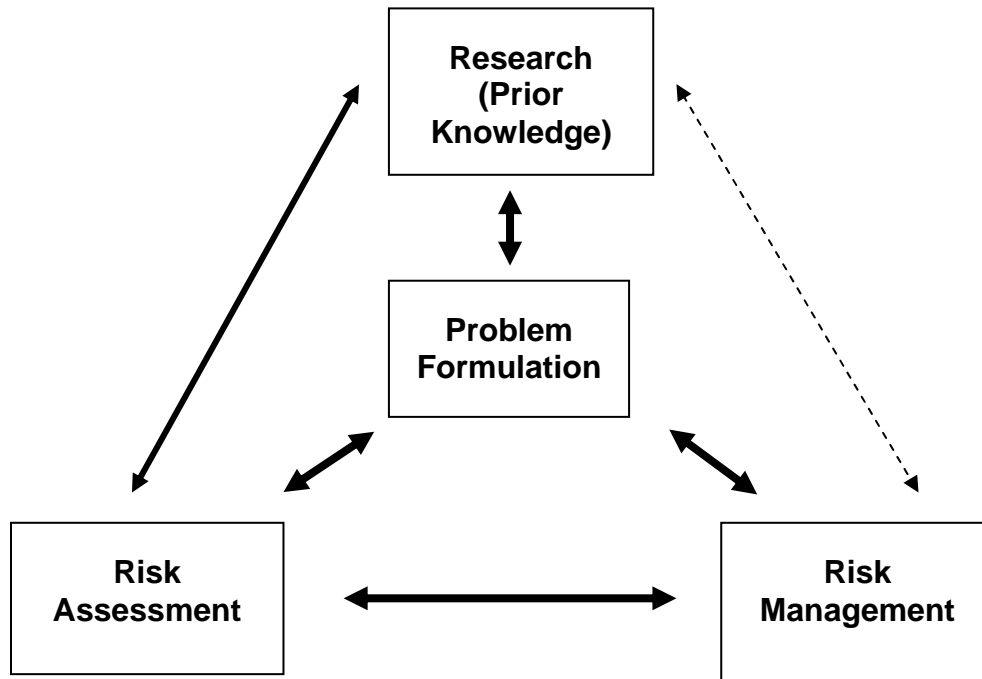
The first risk analysis paradigm for public health was proposed by the U.S. National Academy of Sciences (NAS) (1983) and focused on assessing the risk of cancer from chemicals in food. The decision process was divided into three major steps consisting of research, risk assessment, and risk management. The risk assessment process was further divided into hazard identification, dose-response assessment, exposure assessment, and risk characterization. Risk management is the decision-making process involving the consideration of political, social, economic and technical factors with relevant risk assessment information relating to a hazard so as to develop, analyse, select and implement appropriate risk mitigation options. Risk management comprises three elements: risk evaluation, emission and exposure control, and risk monitoring.

In the NAS paradigm, the principal steps were considered to be sequential, with the decision process commencing with research and concluding with the decision. A drawback of this sequential concept is an absence of the recognition of the influence that the risk analysis might have on data collection, or of the impact political, social, and economic objectives may have on the need to identify the hazard.

More recent examinations of risk assessment/analysis methodology have paid much closer attention to the influence of risk management on the risk assessment process (NAS, 1994; NRC, 1996; Presidential Commission, 1997; Renwick et al., 2003). Rather than insist that management be insulated from the risk assessment process for the sake of preserving scientific objectivity, it is acknowledged that risk management should interact with risk assessment for the scope of the analysis, particularly in problem formulation. The focus on this interaction leads to the notion that the relationship between risk assessment and risk management is an interactive, often iterative, and circular process (see Figure 2.1).

As a general rule, formal risk assessments are preceded by preliminary risk assessments. These are usually subjective and informal and may be initiated from inside or outside the risk assessment and scientific communities. A key consideration of these preliminary risk assessments is whether or not a formal risk assessment is necessary. The transition process from preliminary assessments to formal risk assessments has been described as problem formulation (Renwick et al., 2003). It is an iterative process that facilitates the critical interface between risk assessment and risk management. Risk communication, with stakeholder involvement, is particularly essential during the problem formulation.

1



2

3

4

5

6

**Figure 2.1 Interactions of risk assessment with risk management**

7

8

9

10

11

12

13

14

15

As the risk analysis paradigm evolved, the need for risk communication as an integral part was recognized (see Figure 2.2). Risk communication is not only the interactive exchange of information and opinions among risk assessors and risk managers, but necessarily includes all interested parties. The issues of risk, risk-related factors, and risk perceptions should involve interactive exchange throughout the risk assessment process. The communication of the results of the risk assessment and the bases of the risk management decisions demands transparency and appreciation for the uncertainties involved.



16

17

18

19

20

**Figure 2.2 The risk analysis paradigm**

## 2.3 Motivations and Considerations for Producing a Formal Risk Assessment

There are several different reasons for preparing a formal risk assessment. The relative importance of these different motivations may influence the scope or the methodology used.

### 2.3.1 Transparency and Justification

A major function of formal risk assessment is to serve as a transparent justification of a public health decision, whereby each step and assumption is clearly described. A key reason for undertaking such an assessment is to separate clearly scientific knowledge from values. Formal risk assessments are almost always performed for notable public health issues where there is a wider interest in the political, social and economic consequences of such an assessment. Identifying the public health objectives before the technical analysis will allow participation in the debate of the other issues involved without necessarily requiring involvement in the scientific discussion. There may be areas of a risk assessment that can be obscure to someone not privy to their development. As a result, transparency is often audience-dependent, relative to the level of comprehension and involvement. Since less can be taken for granted, the extent of the explanation required will increase as the audience broadens and their level of interest and sophistication increases. Producing records of an assessment with varying degrees of technical detail may be a useful objective.

The World Trade Organization (WTO), under the Agreement on the Application of Sanitary and Phytosanitary Measures (SBS Agreement) has recognized the importance of harmonized science-based risk assessments. WTO specifically has cited the standards, guidelines, and recommendations of the Codex Alimentarius Commission (Codex) as reflecting international consensus regarding the requirements to protect human health from foodborne hazards. Codex has formally adopted the risk analysis paradigm in its decision making. Other organizations have also adopted this paradigm (European Commission, 2000).

### 2.3.2 Public Health and Individual Health

A public health risk assessment is concerned with a population. The behaviour, environment, or biological characteristics will vary among individuals in the population of concern. A risk assessment may need to describe or model these individual characteristics to produce a prediction of what may be expected to happen in the population. Specifying the population with which the risk assessment is concerned may be an important part of the problem formulation. In a public policy setting, the population will generally be defined by the risk managers, often in view of social, economic and other considerations.

### 2.3.3 Quantification and Computation

Public health issues often involve matters of degree, particularly in regards to level of exposure and risk and may be defined by measures of quantity or statistical rates. If an uncertainty

1 analysis is conducted, knowledge may be quantified as a matter of degree. Although judging  
2 matters of degree does not require the use of numbers, communication of degree does.  
3 Quantitative risk assessment approaches, including DRM, can be valuable in providing  
4 information to address these issues.

5  
6 Formal risk assessments often involve the interaction of multiple quantitative measures that may  
7 lead to extensive and complicated calculations. Particularly in DRM, mathematical and  
8 statistical considerations are often complex. While computers can carry out these calculations  
9 more accurately and quickly, knowledge of the scientific basis and experience with the  
10 applications of DRM are essential in order to avoid misinterpreting and incorrectly  
11 communicating the outcomes.

#### 12 13 **2.3.4 Cost of Assessment**

14  
15 Risk assessments take time and effort to develop. The time and effort required will increase with  
16 the complexity of the problem and often with the degree of transparency that is required. The  
17 level of scientific detail addressed by the models and the level of documentation needed may  
18 vary with the nature and magnitude of the motivations for producing the risk assessment in the  
19 first place. In order to tailor the risk assessment to the decision problem, it may be desirable to  
20 develop the risk assessment by an iterative process that commences with the simplest possible  
21 statement of the problem and becomes more complicated as the risk assessment is developed.

### 22 23 **2.4 Risk Assessment**

24  
25 The risk assessment paradigm, incorporating problem formulation, is illustrated in Figure 2.3  
26 (based on Renwick et al., 2003).

#### 27 28 **2.4.1 Problem Formulation**

29  
30 Problem formulation is the initial phase in a risk assessment that determines if a detailed risk  
31 assessment is necessary and, if so, possible. Further it serves as the transition from an informal  
32 risk assessment to a formal risk assessment. Problem formulation requires at least some  
33 preliminary consideration of the hazard identification, hazard characterization, and exposure  
34 assessment and usually proceeds in iterative stages. The output is a plan for the risk assessment  
35 process, which can be changed as the risk assessment progresses.

##### 36 37 **2.4.1.1 Defining the Question**

38  
39 Among additional considerations are those that address who should be involved in the risk  
40 assessment and risk management processes. The transparency of a risk assessment will depend  
41 on how well these are described. It is not necessary to establish beyond all doubt that there is a  
42 cause-effect relationship in order to conduct a risk assessment. The suspicion that there may be a  
43 relationship is sufficient. The consideration of the evidence for or against the supposition is  
44 often an integral part of the analysis. Identifying the problem may be politically controversial.  
45 That is, it may constitute a risk management issue that must be resolved before the risk  
46 assessment may be used as the justification for a decision. Non-scientific controversy may be

1 diverted from the risk assessment by separating the valuation of the effect from the risk  
2 assessment per se (e.g., the risk assessment may be used as part of a cost-benefit analysis, but the  
3 cost-benefit analysis is not part of the risk assessment). Predicting the occurrence of an event is  
4 not part of an expression of the level of public health concern. However, suggesting that the  
5 problem is big enough to merit a formal risk assessment does imply that the risk may be of some  
6 significance.

7

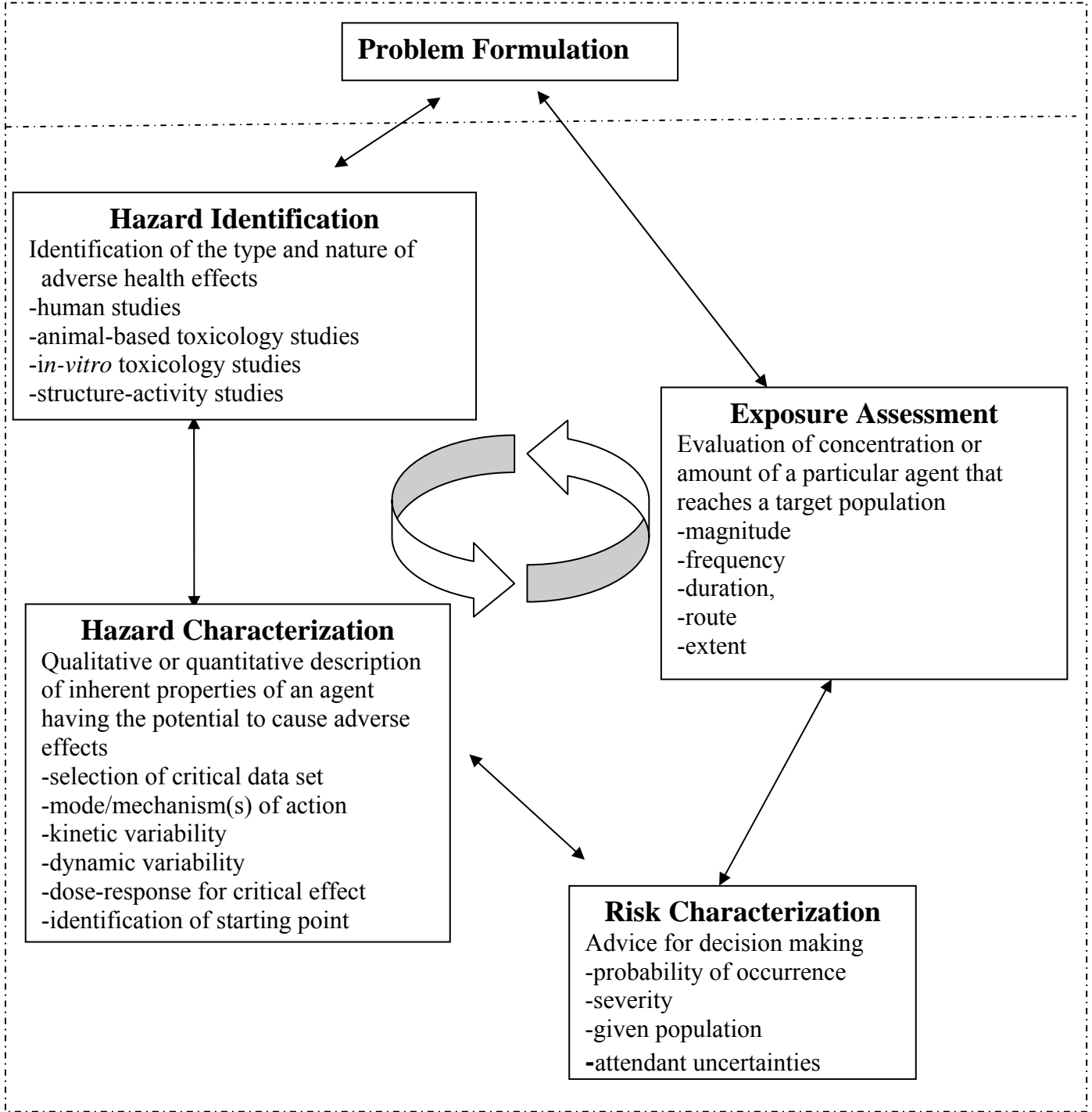
8

#### 2.4.1.2 Prior Knowledge

9

10 Organizing information regarding public health issues that may involve many details and  
11 complex cause-effect relationships may benefit from the methodical collection and evaluation of  
12 prior knowledge of the agent, exposure to the agent, and on possible biological effect(s) resulting  
13 from exposure to the agent. This is essential for determining the feasibility of a detailed  
14 assessment. Prior knowledge is also important for prioritizing and directing the risk assessment.  
15 Organization of information may also instigate and support specialization – different experts may  
16 produce or oversee different parts of the risk assessment. This information may in turn influence  
17 the conception of the problem (where management specifies the objective on the analysis) and  
18 also may influence additional research that may be needed.

1  
2



3  
4  
5

**Figure 2.3 Risk assessment**

### 2.4.1.3 Desired Outcomes

The desired outcomes of the problem formulation are

- explicit questions to be answered in the risk characterization to meet the needs of the risk manager,
- determination of the resources that are needed and available, and
- time frame for completing assessment.

### 2.4.2 Risk Assessment Outcomes

The advice to risk managers that is formulated in the risk characterization may be qualitative or quantitative.

Quantitative advice includes

- health based guidance values,
- estimates of the risks at different levels of exposure,
- exposure-based estimates used with low levels of exposure, and
- risks at minimum and maximum intakes (e.g., nutrients).

Qualitative advice includes

- statements/evidence that the agent is of no toxicological concern due to the absence of toxicity even at high exposure levels (e.g., ADI not specified),
- statements/evidence that the agent is safe in the context of specified use, and
- recommendations for avoidance, minimization, or reduction of exposure.

Risk characterization should include all key assumptions and a clear explanation of the uncertainties in the risk assessment. It should also include information on susceptible sub-populations, including those with greater potential exposure and/or specific physiological conditions or genetic factors. The advice to risk managers can be in the form of a comparison of the relative risks among risk management options.

The risk assessment that is produced is followed by either a risk-management decision or a request for further analysis, which may influence the further research that is conducted. In one sense, the risk assessment process may never end. However, from a risk management standpoint, there is usually some imperative and timeline that concludes the process. Therefore, in another sense, the risk assessment ends when the risk management decision is made. The record produced by a risk assessment stands as a justification for a decision at the time the decision is made. However, with additional information, such as that which can reduce the uncertainties identified in the risk assessment, the risk assessment/analysis may be reopened.

## 1 **3.0 DOSE-RESPONSE MODELLING – BASIC CONCEPTS**

### 2 **3.1 Introduction**

3 Toxicology is the science of identifying and quantifying harmful or adverse effects of chemical  
4 and physical agents in the human environment. This can be accomplished by observations in  
5 humans (i.e., epidemiology and clinical studies), experimental studies using animal models (in  
6 vivo bioassays) or cellular and molecular studies. All these approaches have firmly established  
7 the principle of DR. Accordingly, DR toxicities of chemicals can be and have been expressed  
8 quantitatively (e.g., the LD50). However, scientific data alone is not sufficient to make a  
9 decision regarding the potential toxicity of chemicals and agents that humans encounter; it is the  
10 analysis and interpretation of these data that lead to a scientifically supported decision regarding  
11 potential health effects. Many analytical processes have been developed to address the evaluation  
12 of the toxicities of chemicals ranging from very simple approaches based solely upon the  
13 identification of the possibility of a hazard (NTP, 2002; USEPA, 2003; Cogliano et al., 2004) to  
14 much more complicated approaches incorporating biological mechanisms, complicated  
15 mathematical models, bioavailability in humans and direct predictions of chemically-induced  
16 changes in disease incidence in the affected human population (Portier and Kohn, 1996; Kim et  
17 al., 2002). All of these methods have two basic steps in common: analysis of the DR  
18 information and implementation of the results of that analysis to formulate a conclusion. The  
19 combined two-step approach will be referred to as DRM.

20 This Chapter describes the elements that embody DRM. Most of the information presented is  
21 found in more extensive detail in other chapters of this guidance document. This Chapter sets  
22 the stage for discussion of dose/exposure-response modelling by briefly answering the questions:  
23 What is dose?, What is response?, and What is modelling?. It then goes on to introduce the  
24 reader to the types of data and information that may have an impact on the development of DRM  
25 models.

### 26 **3.2 What is Dose?**

27 It is critical when performing DR analyses to have a clear concept of what is meant by “dose”  
28 and how it applies to the response. There are three basic types of “dose” that arise from scientific  
29 investigations: the administered or external dose, the internal (absorbed) dose, and the tissue or  
30 target dose. External dose denotes the amount of an agent or chemical administered to an  
31 experimental animal or human in a controlled experimental setting by some specific route at  
32 some specific frequency. In the terminology used by the Joint WHO/FAO Expert Committee on  
33 Food Additives (JECFA), intake refers to external dose. Internal dose is the amount determined  
34 by toxicokinetics that is systemically available. It is a consequence of absorption, distribution,  
35 biotransformations, and excretion of the chemical. The tissue dose is the amount that is  
36 distributed to and accumulated in a specific tissue of interest. The three are, of course, related  
37 and each can be used to express DR.

38 Two other parameters are important: the dose frequency and duration of dosing. Dosing can be  
39 acute, subchronic or chronic. For simplicity, the term dose, in DRM will be used as an inclusive  
40 term referring to all three forms of dose described above. In general, units of dose should reflect

1 the magnitude, frequency and duration over which it applies. Dose can be expressed in a  
2 multitude of metrics. Some of these metrics include micromolar concentration, daily intake (e.g.,  
3 ng/kg/day), total body burden (e.g., ng/kg), body burden averaged over a given period of time, or  
4 tissue concentration.

5 For humans, where dosing of xenobiotics is not intentional, the term exposure is used for the  
6 external dose. In epidemiology studies, exposure is rarely known and best estimates are made  
7 using several assumptions and/or biomonitoring of tissue (usually blood) concentrations at very  
8 few time points, often many years after what may be believed to be the period of first/highest  
9 exposure. Sometimes, when laboratory animals are used for DRM, the dose used in the animal  
10 study is transformed to an equivalent human exposure prior to modelling. Exposure assessment  
11 is the qualitative and/or quantitative evaluation of the likely intake of chemical agents via food,  
12 as well as exposure from other sources, if relevant (WHO, 1997). In this situation, models of  
13 exposure linked to response data may be used to develop a DR model. However, limited  
14 knowledge of the events controlling absorption and tissue distribution (especially in humans at  
15 low levels of exposure), biotransformation, and excretion, and the other molecular and  
16 biochemical processes that ultimately lead to particular responses contribute to the uncertainty in  
17 these analyses.

### 18 **3.3 What is Response?**

19 Response, in this context, generally relates to an observation or effect seen in a laboratory cell  
20 culture, an animal or a human following exposure. These endpoints cover a broad range of  
21 observations, from early responses such as biochemical alterations to more complicated  
22 responses such as cancer and developmental defects. Responses can be either adaptive or adverse  
23 (e.g., Williams and Iatropoulos, 2002). The latter are defined as a change in the morphology,  
24 physiology, growth, development, reproduction or life span of an organism system or  
25 (sub)population that results in an impairment of functional capacity, an impairment of the  
26 capacity to compensate for additional stress or an increase in susceptibility to other influences.  
27 These are critical responses that are likely to underlie an adverse health effect in humans. The  
28 responses are sometimes species- and/or tissue-specific and have different degrees of variation  
29 across individuals. Nevertheless, there is some commonality across species and there are known  
30 linkages between some responses (e.g., DNA damage is a precursor for mutations). DRM can  
31 address each response, provide insight into their quantitative similarity across species and tissues,  
32 and link responses in a mechanistically reasonable manner.

33 Response is generally considered to vary across experimental units (animals, humans, cell  
34 cultures) in the same dose group in a random fashion. This random variation is usually assumed  
35 to follow some statistical distribution describing the frequency of any given response for a  
36 population. In general, statistical distributions are characterized by their central tendency  
37 (usually the mean or average value) and their effective range (usually the standard deviation).  
38 Most responses of interest in the context of DR assessment fall into one of four basic categories:  
39 quantal responses, counts, continuous measures or ordered categorical measures. Quantal  
40 responses generally relate to the number of experimental units responding in a given period of  
41 time (e.g., the proportion of animals with a tumor in a cancer bioassay), count data generally  
42 relates to a discrete number of items measured in a single experimental unit (e.g., number of  
43 papillomas on the skin), continuous measures generally take on any value in a defined range (e.g.,

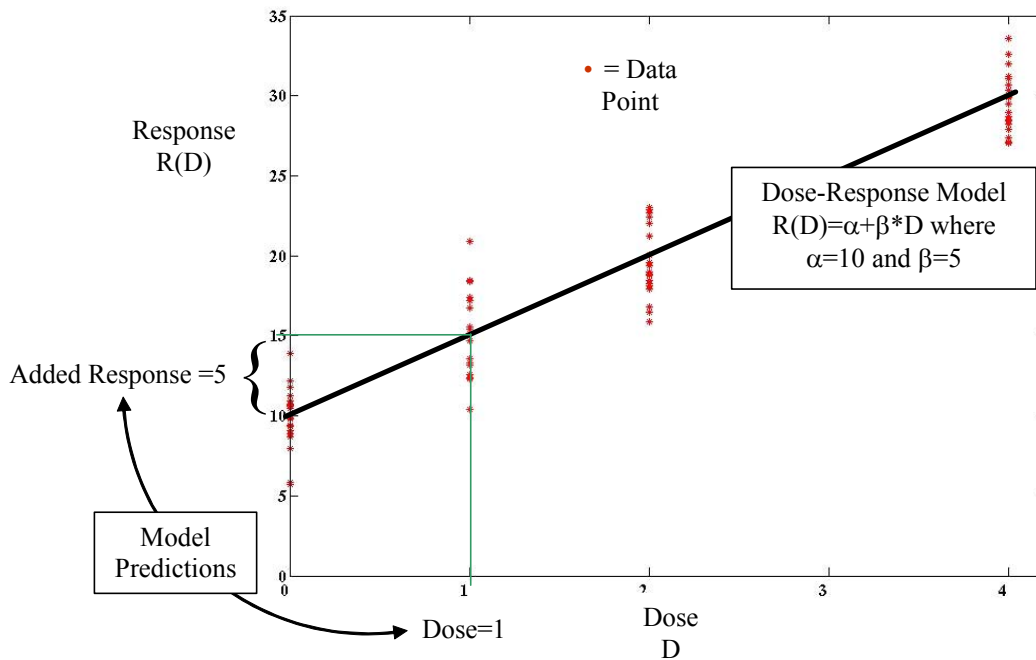
1 body weight), and ordinal categorical measures generally take on one value from a small set of  
 2 ordered values (e.g., tumour severity grades). Sometimes it is useful to convert continuous data  
 3 into proportions (e.g., number of animals outside a clinically relevant range for an immune  
 4 system marker) or categories (e.g., measured degree of liver necrosis converted to minimal,  
 5 moderate, or extensive).

6 For each of these different data types, there will be some differences in how they will be handled  
 7 for DRM; but as a general rule, the goal of DRM is to describe the mean and variance of the  
 8 response as a function of exposure and/or time.

### 9 3.4 What is a Model?

10 DR models are mathematical models used to characterize the relationship between dose and  
 11 response for a given set of scientific data. Mathematical models consist of three basic  
 12 components; assumptions used to derive the model, a functional form for the model and  
 13 parameters that are components of the functional form. For example, the simplest DR model is a  
 14 linear model to describe a continuous response (see Figure 3.1).

15



16

17 **Figure 3.1 Dose-response illustration displaying a linear model fit to continuous data for**  
 18 **which prediction of the dose associated with an added response of 5 units (not**  
 19 **designated) is a dose of 1 unit (not designated)**

1 For this model, the key components are

- 2 • assumptions: mean response is proportional to dose,
- 3 • functional form:  $R(D)=\alpha+\beta\cdot D$  where  $R(D)$  is the mean response as a function of dose
- 4 denoted  $D$ , and
- 5 • parameters:  $\alpha$  is a parameter describing the mean response in the control (unexposed)
- 6 group and  $\beta$  is a parameter describing the mean change in response per unit dose.

7 DR models range from very simple models, such as the linear model described above, to  
8 extremely complicated models for which the eventual functional form cannot easily be expressed  
9 as a single equation (e.g., biologically-based DR models). Models can also be linked, meaning  
10 that one model could describe part of the DR process while another describes the remainder of  
11 the process. For example, in most cases for chemical carcinogenesis, tissue concentration is more  
12 closely linked to cancer risk than administered dose. Given data on dose, tissue concentration  
13 and tumor response, one can use a toxicokinetic model to relate dose to tissue concentration and  
14 use a multistage cancer model to relate tissue concentration to response. The two models  
15 combined are needed to describe the DR relationship.

16 DR models may incorporate other information into the model form. Age and time-on-study are  
17 commonly used in dose-response modelling, but other factors such as species/strain/human  
18 ethnicity, gender, body weight, etc. have also been used to expand the utility of DR models.

### 19 **3.5 What is Dose-Response Modelling?**

20 DRM can be described by six basic steps with a variety of options at each step (Table 3.1). The  
21 first four steps are aimed at the analysis of the data available for DRM, which will be referred to  
22 as DR analysis. DR analysis provides the linkage of a model to DR data for the purposes of  
23 predicting response to a given dose or predicting dose from a given response. The last two steps  
24 deal with implementation and evaluation of the analysis results.

25

26

27

28

29

30

31

32

33

1

**TABLE 3.1: Basic steps in dose-response modelling**

Step	Description	Options	Section Link
1. Data selection	Determine the response to be modeled and select appropriate data	Endpoint, quality, sample size, utility, availability	5.1
2. Model selection	Choose the type of model to be applied to the data	Endpoint, data availability, purpose	5.2.1
3. Statistical linkage	Assume what statistical distributions describe the response	Endpoint, data type, model choice, software availability	5.2.3
4. Parameter estimation	Combine the first three steps in an appropriate computer program to obtain estimates of the model parameters	Linkage function, software availability, variance	5.3, 5.10
5. Implementation	Use the estimated model parameters and the model formula to predict response/dose as needed	Outputs, target selection, model predictions, BMD, Direct extrapolation	5.9, 5.1.2, 5.1.3, 5.1.4, 5.1.5
6. Evaluation	Examine the sensitivity of the resulting predictions to the assumptions used in the analysis.	Model comparison, uncertainty	5.4, 5.7

2 Step 1 involves selection of appropriate data for DRM. The type of data available can have a  
3 marked impact on the complexity of the model that can be used. For example, while two points  
4 can be used to identify the slope of a line, it takes at least three points to identify the shape of a  
5 more complex dose-response relationship (e.g., straight line versus two connected lines). The

1 issue of whether there is enough data to support a given model is quite complex (Portier, 1994)  
2 and is discussed in greater detail elsewhere. In general, the data can restrict the type of model  
3 that can be used.

4 The second step is then to choose an appropriate model. Many choices exist for modelling dose-  
5 response data and examples of some of the possible choices are presented in Chapter 5. These  
6 models have been generally divided into two categories; empirical versus biologically-based  
7 models. Empirical models generally refer to functional forms for which there is limited  
8 mechanistic justification (e.g., the linear model above). Most of the DRM that has been done to  
9 date has focused on the use of empirical models. Biologically-based models generally have  
10 functional forms that are derived from some basic principles about the onset and progression of  
11 disease in a biological system. These models are generally functionally complicated and require  
12 that experience in mathematics, statistics and computer science be linked to experience with  
13 biological mechanisms. Mechanistic models also generally have greater data needs than do  
14 empirical models.

15 The third step requires the choice of a statistical linkage between the data and the model. The  
16 most common linkage method is to assume a statistical distribution for the response and use that  
17 distribution to derive a mathematical function describing the quality of the fit of the model to the  
18 data. However, a considerable amount of dose-response modelling has been done by simpler  
19 linkage functions such as drawing a straight line through the data points. The advantage of  
20 choosing a formal statistical linkage is the ability to test hypotheses and derive confidence  
21 intervals for model predictions.

22 In DRM, fitting the model to the data is the fourth step. Since the primary components of a  
23 model are the parameters that define the model, curve fitting simply involves choosing values for  
24 the parameters in the model. If a formal statistical linkage has been developed for linking the  
25 data to the model, then the parameters are chosen such that they “optimize” the value of the  
26 linkage function. For example, a common choice is to link the data to the model using the  
27 squared distance, denoted  $[R(d_i) - o_{ij}]^2$ , between the predicted value from the model, denoted  $R(d_i)$ ,  
28 and the observed value, denoted  $o_{ij}$ . These squared differences can be summed across all data  
29 points and model parameters are chosen to minimize this sum; this is the common least-squares  
30 algorithm. Simpler methods, such as drawing a line through the data points with a ruler on graph  
31 paper, can also be used to estimate model parameters. For example, by drawing a line through  
32 the data points, the parameters in the linear model can be estimated directly since the value of  $\alpha$   
33 can be estimated as the point where the line crosses the y-axis (zero dose) and the  $\beta$  can be  
34 estimated by calculating the slope of the drawn line. Formal optimization is a better choice for  
35 modelling than ad hoc procedures and should be used whenever possible.

36 The fifth step in DRM is to make the inferences necessary to develop measures to protect public  
37 health. In its simplest form, a DR model allows the prediction of response if the dose is known  
38 and the calculation of the dose if the aim is to target a specific level of response. In addition,  
39 implementation of the DR analysis (Steps 1-4) also encompasses the extrapolation of results  
40 from the specific responses seen for the experiment being modeled to other exposure scenarios  
41 and other doses. This step can also involve an extrapolation from a laboratory species to humans.  
42 Usually, when making a prediction, the emphasis is on the change in response seen in the treated  
43 animals compared with the response seen in the controls. The different types of data (quantal,

1 count, continuous, categorical) require different methods for predicting changes in response  
 2 beyond the normal response. In general, the targets used for additional response fall into the  
 3 categories of added response (simply subtract control response), relative response (fold change  
 4 relative to control response) and extra response (added response scaled to range from zero to the  
 5 maximum possible response). Each of these choices can impact the final decision so care should  
 6 be taken to understand why a specific choice is made. Figure 3.1 illustrates some of the basic  
 7 components of DRM for the simple linear model case and added response.

8 Measures used by public health agencies to prevent excess exposure to a hazardous agent  
 9 generally either fall into the categories of direct banning or limiting exposure. DRM could  
 10 inform both choices although its major impact is in the area of limiting exposure. Several  
 11 methods have been proposed on how to use DRM in this context. The simplest is to use the  
 12 predicted model to find the dose associated with a negligible (e.g., 1 in a million) or zero  
 13 response over control. In general, this results in extrapolation far beyond the range of the data,  
 14 which creates a great deal of uncertainty. A second approach is to use the DR model to identify a  
 15 dose with a known response at or slightly below the observable range (the limit of scientific  
 16 certainty) and use other models to get into a range where the response is assumed to be virtually  
 17 unchanged relative to control response. In this approach, a functional model structure can be  
 18 used, such as a straight line, or something simpler, such as uncertainty factors (UFs), to identify a  
 19 safe level of exposure. All of these options are discussed in Chapter 5.

20 The basic steps in DRM shown in Table 3.1 can be repeated to consider other options in the  
 21 process to understand the impact of choices on the predictions from DRM. This final step (Step 6)  
 22 in DRM is aimed at understanding the sensitivity of the analysis to specific choices and to judge  
 23 the overall quality of the final predictions. The simplest way to evaluate sensitivity is by  
 24 considering several choices and determining if the results dramatically change. Depending on the  
 25 degree of difference between choices, there could be value in performing a formal analysis of the  
 26 quality of the fit of the model to the data. Other methods can also be used to assess the impact of  
 27 choices used in the modelling on the eventual outcome, such as uncertainty analysis, Bayesian  
 28 mixing and other methods. In some cases, Step 6 is performed before Step 5 with a focus on the  
 29 assumptions used in the DR analysis and/or after Step 5 with a focus on the assumptions used for  
 30 implementation. These are described in Chapter 5.

### 31 **3.5 Risk versus Safety in DRM**

32 Risk as used in this discussion is the direct estimation of the likelihood or degree of an event or  
 33 its prevalence in a human population as a function of exposure. Given sufficient data in humans  
 34 in the range of exposure where there is concern, it is possible to obtain scientifically supported  
 35 estimates of risk. In most cases, the data used to develop DR models, are not from studies in  
 36 humans in the range of exposures humans generally encounter. The most common type of data  
 37 used for DRM comes from experiments in laboratory animals, generally at administered doses  
 38 significantly exceeding the exposures humans encounter. Even when human data is available and  
 39 suitable for dose-response analysis, it is generally from select populations, such as workers in  
 40 occupational settings, whose exposures differ from the general population. Thus, in many cases,  
 41 DR analyses need to be extrapolated from an observable region where scientific support is  
 42 available to a region where scientific support is weaker or non-existent. For DR analyses based  
 43 on human studies, the extrapolation is generally a downward extrapolation to different exposure

1 levels but can also include extrapolations to different life stages (e.g., fetus, child) or different  
2 populations with different environmental factors that might affect exposure (e.g., dietary  
3 differences). For DR analyses based upon laboratory data using animals, there is the additional  
4 problem of extrapolating to humans.

5 Most of the methods used to implement (Step 5) the results of a DR analysis are addressing these  
6 extrapolation issues. The methods that have been used for extrapolation are diverse and  
7 sometimes contentious, with different countries, and even different agencies within a given  
8 country, using different approaches. The strategies used for extrapolation basically fall into two  
9 categories: those aimed at using estimates of risk for exposures outside of the range of the data  
10 used in the DR analysis and those aimed at establishing safety without using an estimate of risk.

11 Estimates of risk and the dose associated with that risk generally require extrapolation from the  
12 responses and doses of the data to a lower dose range. These extrapolations can be done using  
13 the model (Step 2) that was fit (Step 4) to the data (direct estimation) or a different model,  
14 usually a line, extending from the lowest dose to a point of zero risk. The latter approach is  
15 generally envisioned to be conservative, assuming that the true risk is less than would be  
16 estimated by this second model at all doses below the dose where scientific support is clear. In  
17 contrast, methods used to establish safety for a given dose without presenting an estimate of risk  
18 rely upon the concept that a dose which is sufficiently distant from the lowest dose associated  
19 with the observable range will be safe. This is generally done using UFs that have been  
20 developed over years of experience, although in some cases where the general human exposure  
21 is estimated, the relative difference between the estimated exposure and the dose at the lowest  
22 edge of scientific support is used (margin of exposure).

23 Regardless of how dose-DR analysis, additional methods are employed to extrapolate to humans.  
24 These methods are also varied, ranging from the use of additional UFs factors to more  
25 complicated modelling schemes based upon difference in toxicokinetics and toxicodynamics  
26 between humans and animals.

27 The term “risk assessment” is generally used to describe the entire process of making a public  
28 health decision regarding a specific chemical or agent. However, risk assessment can be defined  
29 further to differentiate between analyses aimed at establishing safety (as defined above) and  
30 analyses aimed at estimating risks. In this case, “safety assessment” would refer to the decision  
31 process aimed at establishing safety whereas “risk assessment” would refer to assessments aimed  
32 at estimating risks which are part of a larger decision process. Safety assessments are more often  
33 used in cases where exposure can be controlled, such as for food additives and residues of  
34 pesticides and veterinary drugs in foods.

### 35 **3.7 Summary**

36 DRM, as used for informing public health decisions about chemical exposures, is a six-step  
37 process. The first four steps constitute DR analysis and relate to the process through which a  
38 mathematical description of the data is obtained in order to evaluate predicted responses for  
39 known doses or to obtain dose estimates when a chosen response is of interest. The fifth step  
40 involves the implementation of the results of the DR analysis for the purposes of guiding a public  
41 health decision. The final step, which can optionally be applied earlier in DRM, involves an

1 assessment of the quality of the DR analysis and the sensitivity of model predictions to the  
2 assumptions used in the analysis. DRM, because it involves a large number of choices based  
3 upon scientific experience, can take on many different forms and be used in many different ways.  
4 The remaining chapters of this report will focus on the range of choices available for each step in  
5 the process and some guidance to be used to make these choices.

## 4. DOSE-RESPONSE MODELLING – WHY AND WHEN TO USE

DRM is a major part of the hazard characterization within the risk assessment paradigm and has been used in the past for both the characterization of DR relationships observed in animal bioassays as well as for the low dose extrapolation of incidences of adverse effects to the range of human exposure levels. This includes the use of the NOAEL for deriving acceptable guidance levels such as the ADI.

### 4.1 Historical Perspectives

It has always been a challenge to extrapolate from experimental animal bioassays to potential effects in humans in order to protect humans from potentially harmful chemical exposures. A variety of approaches has been developed.

The prototype chemical safety assessment uses the ADI methodology which was introduced by Lehman and Fitzhugh (1954), and has come to be widely employed as health based guidance value (WHO, 1987). The ADI was originally devised as a procedure for the regulatory approval of food additives. Since food additives are deliberately added, the process often defines what the regulatory agency is willing to accept as a legal standard of safety. Because “acceptable” is deemed to be an inappropriate term for chemical contaminants, the same methodology is used to derive a TDI. Comparable terms that have been used are provisional-maximum-tolerable-daily-intake (PMTDI) (see WHO, 1987) or a Reference Dose (RfD), (see Barnes and Dourson, 1988). Other similar methods exist for different types of exposures such as for compounds with accumulating properties, e.g., Provisional Maximum Tolerable Weekly Intake (PMTWI) or Provisional Maximum-Tolerable Monthly Intake (see WHO 1987, 2002).

#### 4.1.1 The NOAEL Approach to Acceptable/Tolerable Daily Intake (ADI/TDI)

Calculation of the ADI can be placed into the context of DRM outlined in Table 4.1.

1 **Table 4.1 NOAEL-derived ADI and BMD-derived ADI (Weibull model) as examples of**  
 2 **DRM**  
 3

Step	NOAEL-derived ADI	BMD-Derived ADI
1. Data selection	Better data sets have a large number of doses, sufficient sample sizes, relevant endpoints in a relevant species given appropriate dosing	Same, but a minimum of two exposure groups (required for Weibull model) and significant dose-response trend
2. Model selection	$R(D) = \begin{cases} 0 & \text{if response at dose } D \\ & \text{is not significantly different} \\ & \text{from control response} \\ 1 & \text{if response at dose } D \\ & \text{is significantly different} \\ & \text{from control response} \end{cases}$	<p>Weibull Model</p> $R(D) = \alpha + (1 - \alpha)(1 - e^{-(\beta \times D)^\gamma})$ <p>is one possible class of models to be applied</p>
3. Statistical linkage	Pairwise statistical tests between dose groups and control group	Response is proportion responding and assumed binomially distributed
4. Parameter estimation	$NOAEL = D_{NOAEL}$ <i>where</i> $R(D) = 0$ <i>for all</i> $D \leq D_{NOAEL}$ <i>and</i> $R(D) = 1$ <i>for all</i> $D > D_{NOAEL}$	Find $\alpha$ , $\beta$ and $\gamma$ that maximize binomial-based log-likelihood function
5. Implementation	$ADI = \frac{NOAEL}{UFs}$	<p>Choose an appropriate response, <math>p</math>, in the range of experimental response. Estimate <math>BMD_pL</math>, the 95% lower confidence bound on the <math>BMD_p</math> where</p> $\frac{R(BMD_p) - R(0)}{1 - R(0)} = p.$ <p>Then <math>ADI = \frac{BMD_pL}{UFs}</math></p>

		<p>Model output is presented in the form of</p> <p>a) Benchmark Dose BMDp (BMDx, DMD0x)</p> <p>b) Benchmark Dose BMD<sub>p</sub>L(BMDLx, BMDLox)</p> <p>c) dose-response function R(D)</p> <p>d) extrapolated values of response or dose</p>
6. Evaluation	Possible only at a limited extent, e.g., by examining the sensitivity of the resulting predictions to the use different statistical testing procedures or to variations in the data	Examining the sensitivity of the resulting predictions is possible for model choice or to variations in the data

1  
2  
3 Selecting the data (Step 1) needed to calculate the ADI is similar to choosing the data to be used  
4 for more complicated modelling; the better data sets have an appropriate number of relevant  
5 doses, sufficient sample sizes, and relevant endpoints in a relevant species. The next step in  
6 calculating an ADI is to determine the NOAEL which is the greatest amount of a chemical,  
7 found by experiment or observation, that causes no detectable adverse effect as defined above.  
8 There is a mathematical model (Step 2), statistical linkage (Step 3) and a method of parameter  
9 estimation (Step 4) that describes the identification of the NOAEL. Consider a response model,  
10  $R(D)$ , of the form:

$$11$$

$$12 \quad R(D) = \begin{cases} 0 & \text{if response at dose } D \text{ is not significantly different from control response} \\ 1 & \text{if response at dose } D \text{ is significantly different from control response} \end{cases}$$

13  
14 The statistical linkage (Step 3) between this model and the data is represented by the statistical  
15 test used to determine if a response at any given dose is different from control. In most cases the  
16 statistical test is used to elaborate whether a dose-response relationship exists for which a  
17 modelling approach can be considered. When the NOAEL approach is chosen, the statistical test  
18 is used to decide upon the existence of a statistically significant increase (e.g., at the 5% level)  
19 over background (e.g., the control group) for each dose level separately. The estimation (Step 4)  
20 of the NOAEL is then achieved by choosing the largest dose,  $D_{NOAEL}$ , for which all smaller doses  
21 have  $R(D)=0$  and all larger doses have  $R(D)=1$ . Mathematically, the estimation can be written as:

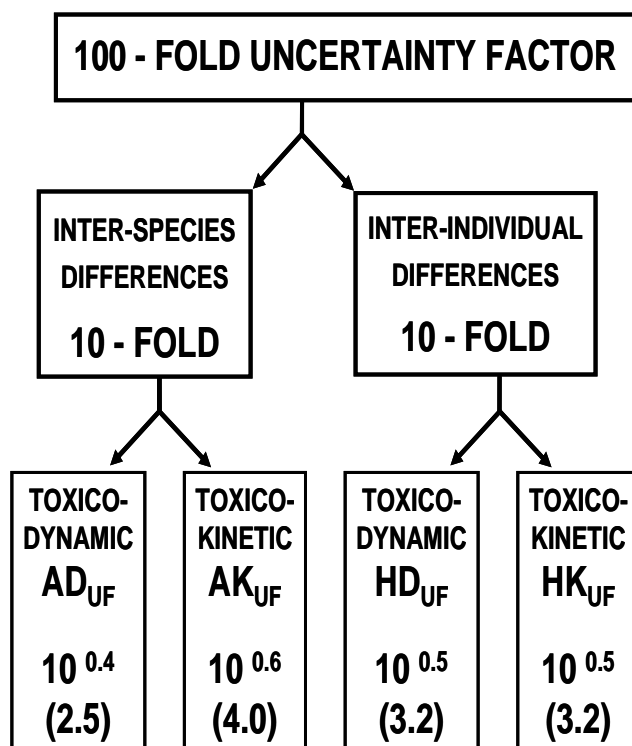
$$22$$

$$23 \quad NOAEL = D_{NOAEL} \text{ where } R(D) = 0 \text{ for all } D \leq D_{NOAEL} \text{ and} \\ R(D) = 1 \text{ for all } D > D_{NOAEL}$$

1 The ADI methodology specifies that an acceptable dose of a chemical may be calculated by  
2 dividing the NOAEL by appropriate uncertainty factors (UFs, also called safety factors). UFs are  
3 default factors used to account for both uncertainty and variability.

4  
5 Historically, an UF of 100-fold has been used to convert the NOAEL from an animal study into a  
6 health-based guidance value (Lehman and Fitzhugh, 1954; Dourson and Stara, 1983; WHO,  
7 1987). Additional UFs may be used to allow for database deficiencies such as the absence of a  
8 chronic study (WHO, 1994). The default 100-fold UF may be seen to represent the product of  
9 two separate 10-fold factors that allow for interspecies differences and human variability (WHO,  
10 1987; Renwick and Lazarus, 1998). The recognition that the original 100-fold UF could be  
11 considered to represent two 10-fold factors allowed some flexibility, because different factors  
12 could be applied to the NOAEL from a study in humans and from a study in animals. The  
13 concept of chemical-specific adjustment factors (CSAFs), (see WHO, 1994, 2001) was  
14 introduced to allow appropriate data on species differences and/or human variability in either  
15 toxicokinetics (fate of the chemical in the body) or toxicodynamics (actions of the chemical on  
16 the body) to modify the relevant default 10-fold uncertainty factor. The strategy used by the  
17 WHO/IPCS involves replacing the original 100-fold UF with CSAFs (WHO, 1994, 2001). This  
18 approach is illustrated in Figure 4.1.

1



Chemical specific data can be used to replace a default uncertainty factor (UF) by an adjustment factor (AF).

A – animal to human; H – human variability; D – toxicodynamics; K - toxicokinetics

2

3

4

5

**Figure 4.1** Subdivision of the 10-fold uncertainty factors to allow for species differences and human variability in toxicokinetics or toxicodynamics (based on WHO, 1994).

6

7

8

9

10

Regardless of the quantities chosen for the UF, the prediction (Step 5) of the ADI from NOAEL-based DRM is given by the equation:

11

12

13

$$ADI = \frac{NOAEL}{UF_s}$$

14

15

16

Step 6 can be extended, both for the NOAEL as well as for the BMD approach, to the evaluation of the sensitivity of the ADI to the assumed values of the UFs.

17

18

19

20

Some scientists have raised concerns regarding the use of the NOAEL to determine an ADI. The greatest of these are that the NOAEL tends to yield lower ADIs for chemicals for which there is more or better data and it is required to be one of the experimental doses (Crump, 1984; Dourson et al., 1985; Kimmel and Gaylor, 1988; Barnes et al., 1995; Slob and Pieters, 1998).

### 4.1.2 Benchmark Dose Approach to Acceptable/Tolerable Daily Intake (ADI/TDI)

As an alternative to the NOAEL approach the Benchmark Dose (BMD) concept was introduced (Crump, 1984; Kimmel and Gaylor, 1988). The BMD method has a number of advantages, including the ability to extrapolate outside the experimental dose range, and responding appropriately to sample size and the associated uncertainty.

In choosing the data (Step 1) for BMD modelling, the same basic considerations apply as for the NOAEL method. In addition, studies showing a graded monotonic response with a significant dose-related trend work best. This is generally true for all DRM analyses.

A generic form of the BMD and Benchmark Dose Lower Confidence Limit (BMDL) is presented in Table 4.1. In Step 5, the designation p, x or ox refers to the Benchmark Rate (BMR). In this document, for example, a variety of response levels such as 1%, 5% and 10% will be discussed.

Choosing a model (Step 2) for the BMD method is dependent upon the types of data available and the characteristics of the response being modelled. More complicated models will require a larger number of dose groups than simpler models and more complicated experimental designs. Several models have been proposed for each type of data. In the USEPA BMD software program a number of routinely used models are cited (<http://cfpub2.epa.gov/ncea/cfm/recordisplay.cfm?deid=20167>). As an example, assuming the availability of data that represent the proportion of animals responding to a given exposure with an adverse effect (e.g. cancer) from each dose group, one model choice could be the Weibull model which has the form:

$$R(D) = \alpha + (1 - \alpha)(1 - e^{-(\beta \times D)^\gamma})$$

where  $\alpha$  is the proportion responding in the unexposed group,  $\beta$  describes the increase in probability of adverse effect per unit dose and  $\gamma$  describes the shape of the dose-response curve (e.g.,  $\gamma > 1$  implies threshold-like behaviour,  $\gamma = 1$  implies linear behaviour).

The statistical linkage (Step 3) between the data and the model can assume a number of different forms as described earlier (Section 3.5). For the Weibull example, it is appropriate to assume the data are binomially distributed. Estimating model parameters (Step 4) for the BMD method also can be based upon a variety of different methods. For the Weibull example, one routinely used approach would be to choose the parameters that maximize the binomial-based log-likelihood.

The concept of the BMD comes from the idea that it is desirable to use a DR model to capture the general pattern of response for all dose-groups in the experimental data set, but there was some dose, the BMD, below which predictions would be tenuous. This BMD can be selected in a number of ways (e.g., Barnes et al., 1995; Murrell et al., 1998), but the most common way is to choose an excess response, the BMR, p, below which there was insufficient support from the

1 data. A common choice for BMR is  $p = 10\%$ . Once the BMR ( $p$ ) is selected, the BMD,  
 2 specifically denoted  $BMD_p$ , is calculated according to the following equation:

$$3 \quad \frac{R(BMD_p) - R(0)}{1 - R(0)} = p$$

5  
 6 The 95% statistical lower bound on the estimated BMD is similar to a NOAEL and this value  
 7 could be substituted for the NOAEL in calculating the ADI (Crump, 1984; Barnes et al., 1995).  
 8 As with all aspects of modelling, many choices exist for calculating confidence bounds and will  
 9 be discussed further in Chapter 5. Having chosen a method for estimating a 95% statistical lower  
 10 bound on  $BMD_p$ , which can be called  $BMD_{pL}$ , the ADI can be calculated as follows:

$$11 \quad ADI = \frac{BMD_{pL}}{UFs}$$

13  
 14 In this calculation, the values of the UFs could be the same as used for the NOAEL or adjusted to  
 15 account for a slightly different interpretation for the  $BMD_{pL}$  relative to the NOAEL (Renwick et  
 16 al., 2003).

17  
 18 Table 4.1 summarizes the implementation in its step 5. This is the step where the model output of  
 19 DRM is described. Three major types of output can be generated:

- 20
- 21 - benchmark doses (BMDs),
- 22 - DR functions, and/or
- 23 - extrapolated risk estimates.
- 24

25 The BMD method includes the determination of the response at a given dose, the dose at a given  
 26 response, and their confidence limits. Using extrapolation of the DRM below the biologically  
 27 observable dose range, the response at specified (lower) dose levels can be estimated as well as  
 28 the dose corresponding to a specific response level.

## 30 **4.2 Points of Consideration**

31  
 32 The use of DRM in general for hazard characterization is possible when a sufficient amount of  
 33 dose response information is available, either from an experimental animal bioassay or from a  
 34 human study (epidemiological study/clinical trial). DRM becomes relevant for risk assessment  
 35 when there is a need to base the starting point for the low dose extrapolation on more dose-  
 36 response information than on that represented by a single dose-response point such as the  
 37 NOAEL, and in particular, when a quantification of the dose-response relationship of the  
 38 observed data is required. DRM may be carried out when biological information and the  
 39 available data are sufficient to extrapolate beyond the observed dose-range.

### 41 **4.2.1 General Aspects of Definition**

42  
 43 The NOAEL is a parameter directly derived from the observed dose-response curve and is  
 44 defined as the highest administered dose where the effect is still not significantly different from

1 that at dose 0 (see Section 4.1). The NOAEL is based on a multiple test procedure performed  
 2 along the dose-response curve. It lacks further statistical properties compared to a parameter of a  
 3 DRM for which a lack of bias and precision would be assessable.

4  
 5 The dependence of the NOAEL on the statistical significance test, however, tends to penalize  
 6 chemicals for which there is more or better data by giving a higher estimate for those chemicals  
 7 with less precise data. This issue can be addressed by using DRM.

8  
 9 The NOAEL approach can be formally considered a dichotomous model where no effect is  
 10 present below and where an expression of the critical effect is present above, see Section 4.1 and  
 11 Table 4.1. In contrast a DRM in general breaks down this sort of dichotomy of effect and no-  
 12 effect levels. The NOAEL approach does not provide quantitative information about risk above  
 13 the ADI. Such information may be provided by DRM

14  
 15 The NOAEL identifies a single dose which is assumed to be without appreciable effect, while  
 16 DRM estimates are based on data from the entire dose-response curve for the critical effect. The  
 17 DRM thus reflects the characteristics of the dose-response curve, particularly in providing  
 18 estimates of slope, and in the case of a regression framework it provides standard error and  
 19 confidence intervals for the model parameters.

#### 20 21 4.2.2 Estimation Procedure

22  
 23 NOAELs are restricted by the set of doses used in the specific studies. An important  
 24 consequence is that the NOAEL may be either below or above the threshold it aims to  
 25 approximate, assuming one exists.

26  
 27 The value of the NOAEL depends strongly on the following characteristics of the study design:

- 28  
29 • Group size. The power to detect a NOAEL at some dose level is directly dependent on  
30 the sample sizes chosen at those dose levels (Gaylor, 1989). The larger the group size,  
31 the greater will be the precision of the NOAEL estimate.
- 32  
33 • Dose spacing. The NOAEL is also influenced by the selection of the dose levels and the  
34 magnitude of effect at higher dose levels. If the doses are widely spaced in relation to  
35 the slope of the dose-response curve, and the dose (LOAEL) above the threshold  
36 produces only a slight effect, then the NOAEL will probably underestimate the  
37 threshold. However, if the doses are more closely spaced then the NOAEL may be at or  
38 above the threshold. Dose placement is a critical factor, and to minimize the probability  
39 of inadequate dose placement the use of multiple dose studies is favourable (Slob et al.,  
40 2005).
- 41  
42 • Sensitivity of measurement of the adverse effect. The more sensitive the method of  
43 detection, the lower will be the NOAEL. Poor or inadequate methods can result in a  
44 higher NOAEL. Studies which are submitted for regulatory purposes should comply  
45 with GLP or be reported in sufficient detail to provide information on the measurement  
46 sensitivity and allow for quality assurance to those undertaking the safety assessment.

1  
2 DRM derived estimates are based on interpolation and do not depend directly on sample size.  
3 DRM can also be used on a study where a NOAEL can not be defined, so in this situation  
4 another study may be unnecessary. A cautionary note is for the potential to "model shop" with  
5 DRM to select those models which produce a predetermined outcome.  
6

7 It should be noted that in comparison to the NOAEL approach, implementation of the DRM  
8 approach leads to BMD doses that are quite similar to NOAELs for the studies in question. The  
9 adoption of DRM does not lead to a wholesale change in the notion of what constitutes  
10 acceptable/tolerable doses, and the continued use of NOAELs (when BMDs cannot be  
11 calculated) is eased because they can be reasonably considered to be approximately comparable.  
12 A DRM, such as the BMD, can be viewed as strengthening the NOAEL by assessing the slope  
13 and the outcome of the assessment could be fairly similar.  
14

15 The outcomes of hazard characterisations based on the NOAELs are not greatly different from  
16 those based on BMDs that represent the lower confidence interval on the dosage giving a 5% or  
17 10% response (Renwick et al., 2003).  
18

### 19 **4.2.3 Uncertainty**

20

21 A modelling approach facilitates both sensitivity and uncertainty analyses. Uncertainty (see  
22 Chapter 5) can be expressed numerically when the doses and responses are linked by a model  
23 (see Table 4.1). Such numerical analyses can also be subject to sensitivity analyses, to test the  
24 contribution of different aspects of the database or of model characteristics to the overall  
25 uncertainty. The uncertainty that arises from aspects of study design, such as dose spacing,  
26 sample size and biological variability on the risk estimates can be assessed in a DRM. Neither  
27 the threshold model of the NOAEL, nor default uncertainty factors are readily amenable to  
28 quantitative estimation of uncertainty or to a sensitivity analysis. Table 4.2 identifies approaches  
29 to uncertainty at each of the six steps of the dose-response modelling scheme of Table 4.1.  
30 When comparing the NOAEL approach with the BMD approach it becomes obvious that DRM  
31 allows the consideration of uncertainty at each step.  
32

33 A disadvantage using the threshold model of the NOAEL for the estimation of a "starting point"  
34 for formulating advice to risk managers is that it is not possible to quantify the degree of  
35 variability and uncertainty that may be present. Advice is limited to comments such as:  
36

- 37 • The NOAEL is assumed to be close to the threshold level when good toxicological  
38 studies with recommended group sizes and narrow increments between dose levels have  
39 been used.  
40
- 41 • NOAELs from studies using a poor toxicological study design such as small group sizes  
42 may be higher than the threshold modelled by the NOAEL approach  
43
- 44 • Studies using wide increments between dose levels may produce NOAELs that are  
45 considerably lower than the threshold modelled by the NOAEL approach (see Renwick  
46 et al., 2003).

1  
2 Most toxicity studies investigate several effects, so care is needed in interpreting DRM. For  
3 example, minimally adverse effects are usually the first to occur as the dose levels increase.  
4 Some of these effects disappear at higher doses when more severe ones become apparent.  
5 Modelling of these less severe effects can lead to spurious results, or to results that are difficult  
6 to interpret because the response does not apparently increase with dose. Applying DRM to  
7 human epidemiological studies and clinical trials introduces further uncertainties such as  
8 confounding, sample size limitations, and human variability.  
9

#### 10 **4.2.4 Study Design**

11  
12 Better study designs (e.g., larger sample sizes or optimally placed dose levels) can result in  
13 reduced uncertainty. In principle, the DRM approach requires various dose groups with different  
14 response levels rather than many replicates (animals) within dose groups.  
15

16 A design optimal for one DRM could limit the use of other DRMs and, in particular, the use of  
17 traditional methods of safety assessment (such as the NOAEL). While standard designs used for  
18 the determination of a NOAEL may not be optimal for DRMs in general and the calculation of a  
19 "starting point". The quantity and quality of data determines how complex the DRM can be.  
20 However, a full characterization of the dose-response relationship may not be possible when only  
21 limited data are available.  
22

23 An important point to bear in mind is that DRM can be used on studies carried out in the past and  
24 based on the traditional design. Some have argued that optimal designs for DRMs may have the  
25 advantage for animal welfare that fewer animals could be used.  
26

27 While DRM provides uncertain estimates when the number of dose groups is too small, both the  
28 determination of the BMD/BMDL and that of the NOAEL may both prove inadequate at  
29 different points when the number of animals per dose group is too small. For example, when the  
30 critical effect is seen in a larger experimental animal, such as the dog, with few animals per dose  
31 group, the NOAEL may be high due to the insensitivity of the test. The BMD/BMDL approach,  
32 however, can be used to evaluate sparse dose-response data, and quantify the inherent  
33 uncertainty. However, even here, where an apparent DR relationship in the data remains, the  
34 BMD/BMDL may also provide very uncertain estimates. Therefore, a typical four-dose study  
35 with a few animals per dose may in practice be unreliable whatever method, NOAEL or BMD, is  
36 applied.  
37

38 DRM reduces the need for more experiments when a small degree of extrapolation is needed, e.g.  
39 when the exposure data are near the human exposure level. Whereas, the NOAEL approach may  
40 then require further experiments, for an example see Allen et al. (1996) where a study on boric  
41 acid in the diet failed to establish a NOAEL whereas the BMD approach could have been applied  
42 thereby avoiding the need for a repeat (see also Section 4.2.2 above). Distributing the total  
43 number of animals over more dose groups must not result in poorer performance despite the  
44 smaller number of animals per dose group as shown by Slob et al., (2005). The example on  
45 developmental risk assessment in rats exposed to boric acid in their diet (Allen et al., 1996)

1 suggests that the BMD approach provides a reasonable basis for appropriately comparing and  
2 combining studies opposed to ad hoc combinations of study results.

3  
4 A major advantage of DRM is the ability to estimate risks at or near the range of exposure levels.  
5 In animal studies it is possible to estimate risk over the full range of doses used. The risk  
6 estimation below the BMD depends on the specification of a low level effect that would typically  
7 be unobservable.

8  
9 Some data could support extrapolation of a fitted model beyond the starting point of a low dose  
10 extrapolation when there are indications that the same toxicological mechanism is still active in  
11 the extrapolation region as in the experimental region of the model fit. However, it may be  
12 difficult to find the mathematical model which is most appropriate for describing the  
13 mechanisms and to estimate the precision and uncertainty of the predictions from that model

#### 14 15 **4.2.5 Biological Information**

16  
17 The NOAEL approach incorporates biological information through the application of 'expert' but  
18 subjective judgement. DRM has the potential for a more 'science rich' analysis through the more  
19 formal quantitative inclusion of factors/covariates into the models, particularly in the case of  
20 human epidemiologic data.

21  
22 Such an approach can lead to more certain estimates centred on a toxicologically-based concept  
23 of estimating the dose-response relationship on the basis of all available biological knowledge  
24 using empirical data and applying statistical inference. More complicated models can be  
25 developed on the basis of toxicokinetics and toxicodynamics.

#### 26 27 **4.2.6 Comparison of Experimental Results**

28  
29 NOAELs derive from an algorithmic analysis of the results of a single experiment. Meta-  
30 analysis on data such as NOAELs across a range of studies or chemical is possible such as when  
31 data are insufficient to build a dose response model but may be limited by the statistical  
32 properties of the NOAEL estimates

33  
34 Estimates derived from DRM, however, enhance the ability to compare different experiments,  
35 effects and compounds using a common framework. The estimates obtained may provide a test  
36 of consistency among different studies which may use different dose levels. DRM methodology  
37 can be used to describe dose-response relationships in different studies (e.g. rat and mouse,  
38 chronic and sub-chronic exposure, healthy and diseased animals) if suitable data sets exist.

39  
40 Rules for combining studies, however, need to be developed. Descriptions of the dose-response  
41 on the same endpoints in different studies may be integrated to provide a cohesive picture of the  
42 chemical's toxicity. The values obtained using DRM may result in estimates for each endpoint  
43 on the basis of biological and functional relevancy.

44

### 1           **4.2.7 Risk Management Perspectives**

2  
3 The potential use of the estimates from DRM can, from a risk management perspective, give an  
4 improved characterization for decision making by

- 5
- 6       • providing information about what happens above the safety level (magnitude and types  
7       of health impacts);
- 8
- 9       • showing risks benefits from different regulatory actions; give the decision maker a  
10       'more-than-one-point' appreciation of the data;
- 11
- 12       • promoting consistency in decisions, if appropriate adjustments are made for differences  
13       in effect, effect level, species, and study design; and
- 14
- 15       • finally, DRM should promote an iterative interaction with risk managers on a  
16       continuous basis
- 17

### 18           **4.3 Implementation Issues**

19  
20 Risk assessment requires in most applications the extrapolation to doses which are lower and,  
21 usually, much lower than the doses used for the analysis of a DRM. The experimental data  
22 obtained in animal experiments are almost always obtained at doses orders of magnitude larger  
23 than those assessed for the risk assessment. DR data from human exposure are rare or also  
24 obtained at higher dose ranges (e.g. in chemical accidents or at occupational exposure). Risk  
25 assessment has to bridge the gaps by extrapolation both from high to low doses and from animals  
26 to humans. There is no simple algorithm for the extrapolation step of the dose-response  
27 assessment. Any extrapolation should use relevant information about those biological processes  
28 operating at the intermediate dose interval, or which can explain interspecies differences. One  
29 approach to fill the gap between high and low doses is to use biological knowledge to determine  
30 the shape of the DRM between the low dose region and the region of the empirical model fit.

31 The fundamental objective of risk assessment, as with modelling generally, is extrapolation.  
32 Experimental data observed under specified conditions are extrapolated to human exposure  
33 conditions that may be similar in some respects and dissimilar in other respects from those in the  
34 experiments. Similarly, epidemiological data observed in one human population exposed to  
35 certain conditions may be extrapolated to other human populations and other exposure  
36 conditions. The models developed as part of a dose-response assessment describe how these  
37 extrapolations are made and provide a basis for evaluating how well the extrapolations are  
38 supported by data.

39  
40 The main strength of the DRM approach is that an explicit response level can be associated with  
41 it. Thus, the dose serving as a starting point for human risk assessment is based more on  
42 toxicological judgement (e.g., choice of BMR), and somewhat less on the statistical  
43 characteristics of the data (which is sometimes more the case in assessing a NOAEL).  
44

1 The analysis of the DR in a dose range where effects are actually observed is the first step of  
2 low-dose extrapolation. The second step would then be the method which determines the risk  
3 estimates in the low dose region to be evaluated as without harm to humans.

4  
5 This two-step approach appears suitable for overcoming deficiencies of earlier concepts where  
6 the ADI was estimated more or less in one step from the empirical dose-response data, e.g., using  
7 the NOAEL or LOAEL together with some default uncertainty factors. The BMD method  
8 emphasizes the application of this two step approach. The DRM, as outlined in this document,  
9 may restrict model predictions to the range of observed dose-response data. It clearly separates  
10 the task of setting an ADI/TDI/RfD from the scientific task of estimating a DR curve.

11  
12 In the case of the BMD there are a number of decisions to be made in applying the method and  
13 determining a BMDL; for example, which mathematical model to use; what degree of  
14 confidence to use in calculating confidence limits; what response level to predetermine as the  
15 benchmark response (e.g., BMR = 1%, 5% or 10% incidence of an effect, or a 5% or 10%  
16 change in a continuous endpoint, such as body weight or red blood cell counts). It is often not  
17 clear what response level BMR can be considered as non-adverse. For example, should a 5%  
18 decrease in red blood cell counts be considered as adverse, or should a smaller (or larger) change  
19 be chosen? Or, should up to a 5% increased incidence in hepatocellular hypertrophy be  
20 considered as acceptable in an animal study, or is a maximum of 10% increase adequate? These  
21 and other choices need additional discussion among toxicologists and medical practitioners.  
22 Although an explicit statement on the BMR is an improvement compared with the generally  
23 unknown response level associated with a NOAEL, choices of a BMR need consensus building.  
24 An important contribution of the BMD approach is that it helps the risk assessor to consider  
25 further the toxicological and statistical aspects of the data, thereby allowing any discussion on  
26 the appropriateness of the BMR to be influenced by toxicological considerations.

27  
28 A feature of DRM is that a large family of models give a good fit to the experimental data, but  
29 the risk estimates obtained by extrapolation to very low doses can be very different and no  
30 further information may be available to discriminate between these models. This means that the  
31 risk assessor must make the risk manager aware of this significant source of model uncertainty  
32 (see Table 4.2.)

33  
34  
35

**Table 4.2 Approaches addressing uncertainty in the process of dose-response modelling and how they are captured by the NOAEL and the BMD method.**

Step	Source	NOAEL-derived	BMD-derived
1. Data selection	Choice of data sets	OK	OK
	exposure/dose response	OK	OK
	Inter-individual variability (population heterogeneity)	not possible	OK
	Confounding (human studies)		
2. Model selection	Choice of model incl. choice of model parameters	not possible	OK
3. Statistical linkage	Choice of distribution	may be possible	OK
4. Parameter estimation	Choice of endpoint	OK	OK
	Statistical error ( standard error, confidence interval)	not possible	OK
5. Implementation	Choice of starting point	not possible	OK
	Choice of method of low dose extrapolation	OK	OK
	using uncertainty factors using a linear extrapolation to dose 0 (control)	not possible	OK
	Extrapolation of the DRM between species between routes of exposure		
6. Evaluation	Sensitivity to model choice	Not possible	OK
	Sensitivity to statistical tests	OK	not necessary
	Sensitivity to variations in the data	OK	OK

1  
2  
3  
4

5

#### 1 4.4 Summary

2  
3 The characterization of DR relationships in animal and human studies has been a major  
4 component of hazard characterization and has been used in the extrapolation of incidences of  
5 adverse effect in the range of human exposure levels. Over the years a variety of methods have  
6 been developed to accommodate such relationships, including the NOAEL and the BMD, as well  
7 as other extrapolation methods for improving extrapolation to low doses. DRM may prove  
8 useful in the risk assessment consideration when a sufficient amount of animal or human dose  
9 response data is available and the uncertainties associated with such an extrapolation are deemed  
10 to be tenable. The standard NOAEL approach, a special case of DRM, identifies a single dose  
11 which is assumed to be without appreciable effect, while DRM estimates in general are based on  
12 data from the entire DR curve for the critical effect. The DRM thus reflects the characteristics of  
13 the DR curve, particularly in providing estimates of slope, and in the case of a regression  
14 framework, it provides standard error and confidence intervals for the model parameters. A  
15 disadvantage of using the NOAEL is that it is not possible to quantify the degree of variability  
16 and uncertainty that may be present, while other DRMs can facilitate the analysis of sensitivity  
17 and/or uncertainty. Consideration of a DRM can optimize study design and clarify the need for  
18 additional studies. The NOAEL approach incorporates biological information through the  
19 application of 'expert' but subjective judgment. DRM has the potential for a more 'science rich'  
20 analysis through the more formal quantitative inclusion (e.g., factors/covariates) into the models.  
21 Estimates derived from DRM enhance the ability to quantitatively compare different experiments,  
22 effects and compounds within a common framework. DRM can enhance the safety assessment  
23 paradigm as well as providing opportunities to consider the likelihood of effects outside the  
24 observable range.  
25

## 1 **5. PRINCIPLES OF DOSE-RESPONSE MODELLING**

### 2 **5.1 Data**

#### 3 **5.1.1 Selection of Data**

4  
5 When considering which data to use from a set of available toxicity studies on a particular  
6 compound, it may not be effective to do a DR analysis for each observed endpoint in each study.  
7 As a first step one may omit studies that have obviously larger NOAELs compared to the other  
8 studies. In this way one may, for example, select for a given type of toxic response (e.g. chronic,  
9 developmental) for the most sensitive species. For a given study, many endpoints may have  
10 been measured. Endpoints not showing a clear DR on visual inspection can be omitted. Then,  
11 based on the toxicological impact together with the apparent magnitude of the response, a  
12 selection of endpoints can be made as candidates for modelling. It would be very helpful if  
13 submitted studies have an annex with plots of the results (in addition to tables) of observed data  
14 points for each endpoint, possibly with a fitted curve to it, to enhance the process of selecting  
15 endpoints.

16  
17 After selecting the potentially relevant endpoints, the question is if each DR dataset actually is  
18 amenable to do a DR analysis. Generally it is desirable to have at least three, but probably four  
19 different doses (including controls). In addition, the associated effect levels need to be different  
20 from each other; it is preferable to have at least three different response levels. However, for  
21 quantal data at least two partial responses would be needed.

22

#### 23 **5.1.2 Data Types**

24  
25 There are various types of response data, and these can be categorized in various ways. The  
26 main distinction relevant for effects is that between quantal and continuous data. Quantal data  
27 relates to an effect which is observed or not in each individual subject (laboratory animal or  
28 human). Hence, for each dose the number of subjects responding out of the number of subjects  
29 available is reported. In continuous data a quantitative measurement is associated with each  
30 individual subject. As an intermediate type of data, ordinal data reflect (ordered) severity  
31 categories, i.e. they are qualitative data but with a rank order (e.g., histopathological data).  
32 When the categories are non-ordered they are called categorical data, but these are rare for  
33 response data. Finally, count data form another class of data (viz. discrete data), but in practice  
34 they often can be treated as continuous data.

35

36 Although the type of data is important for statistical reasons (see Section 5.2.2 on distributions),  
37 the distinction between quantal and continuous data also has a crucial impact on interpretation of  
38 results, and their ensuing use in risk assessment. In the case of quantal dose-response data,  
39 information on the change of incidence with dose is available at one particular degree of effect.  
40 For example, the incidence of cleft palate may increase as dose increases, but under the  
41 categories “no cleft palate – cleft palate”, there is no information about the degree of the effect.

1 In ordinal and continuous data, in contrast, information on both the degree of effect and the  
 2 incidence is available as a function of dose. So, for example, cleft palate might be categorized  
 3 into an ordinal variable with levels no clefting, mild, moderate, and severe clefting, or might be  
 4 quantified in a continuous variable as, for example, the fraction of closure. The relationship  
 5 between the average response and dose gives information about how exposure changes the  
 6 degree of effect. For instance, a plot of (average) red blood cell (RBC) count may show the  
 7 decrease in mean RBC count (i.e., the degree of the effect) as a function of dose. By also  
 8 considering the individual data points, information on the incidence can be derived as well. For  
 9 example, an estimate the fraction of individuals with RBC counts less than some critical value  
 10 can be derived.

11  
 12 When using animals as a model for the human exposure, the observed dose-response information  
 13 is assumed to mimic to some approximation the dose response in humans. It might be argued that  
 14 this assumption is more plausible for degree of effect than for incidence. The problem is that the  
 15 observed dose-incidence relationship from animals largely reflects the variation in the animals  
 16 used, which is highly controlled in a laboratory experiment. Hence, it may not mimic the human  
 17 variation.

## 18 **5.2 Models and Distributions**

### 19 **5.2.1 DR Models**

#### 20 **5.2.1.1 Continuous DR Models**

21  
 22 The models listed in Table 5.1 are some of the equations that may be used to describe the  
 23 relationship between dose and the magnitude of a response on a continuous scale in an individual.  
 24 When combined with a statistical distribution (e.g., normal or lognormal), these equations can  
 25 also be used to describe the relationship between dose and a continuous response in a population,  
 26 where the continuous model corresponds to the central estimate.

27  
 28 DR data are often adjusted by subtracting the (mean) control value from each individual  
 29 observation. However, this procedure does not account for the fact that the background response  
 30 level in the controls is, like the experimental groups, subject to sampling error. A better approach  
 31 is to account for the background response in the model with a parameter that needs to be  
 32 estimated from the data. Among many of the ways this can be done, the following are three of  
 33 the simplest:

- 34
- 35 1.  $y = a + f(D)$
- 36 2.  $y = a \times f(D)$
- 37 3.  $y = f(a+D)$
- 38

39 where  $a$  is the background term and  $f$  may be any DR function. For some assessments there may  
 40 be mechanistic information that makes one form more preferable than another. For example, the  
 41 first form is preferable for modelling an influence that produces the effect independently, the  
 42 second corresponds to the idea of normalizing the response as a fraction of the background

1 response, while the third reflects a contribution from another agent acting by the same  
2 mechanism

### 3 **5.2.1.2 Quantal DR Models**

4  
5 Quantal DR functions describe the relationship between dose and the frequency of a particular  
6 outcome in a population (see Table 5.2). For a group of homogenous or nearly identical  
7 individuals, the relationship between dose and frequency can be described with a step function  
8 where all subjects either respond or fail to respond at any given dose. However, because  
9 variability is ubiquitous in living organisms, quantal dose-response data typically show gradually  
10 increasing incidences with dose. One interpretation of this is that individual subjects differ in  
11 tolerance to the agent, which can be described by a statistical tolerance distribution. Hence, any  
12 cumulative distributing function may be used as a quantal dose-response function. Other models  
13 have been derived from statistical assumptions about how the agent might exert its effect in an  
14 organism, such as the multi-hit gamma model.

15  
16 Background response rates should, just as in the case of continuous data, be accounted for by  
17 incorporating an additional parameter in the dose-response model. The two simplest ways of  
18 doing this are:

$$19 \quad 1. \quad y = a + (1 - a) f(x)$$

$$20 \quad 2. \quad y = f(x+a)$$

21  
22 where  $f(x)$  is any DR function (varying from 0 to 1). As with continuous data, correcting the data  
23 for background-response prior to the dose-response analysis is statistically unsound. The  
24 background response level should be estimated simultaneously with the dose-response model,  
25 and be treated in the same way as the observed responses in the other dose groups.

### 26 **5.2.1.3 Thresholds**

27  
28 The term “threshold” can be used in three different senses. First, it is used in a scientific sense  
29 to indicate a level of exposure at which no effect occurs (e.g., there is a physical stimulus, but  
30 there is no response). Second, a threshold may be thought of as a level at which there may or  
31 may not be an effect, but it is too small to be observed (e.g., a No Observed Adverse Effect  
32 Level). In this case, it is the perceptual limitation of an observer or analyst, rather than the actual  
33 subject of the experiment that is being described. As a third meaning, a “practical threshold” is a  
34 response where the consequences are determined to be trivial and not worth further consideration.  
35

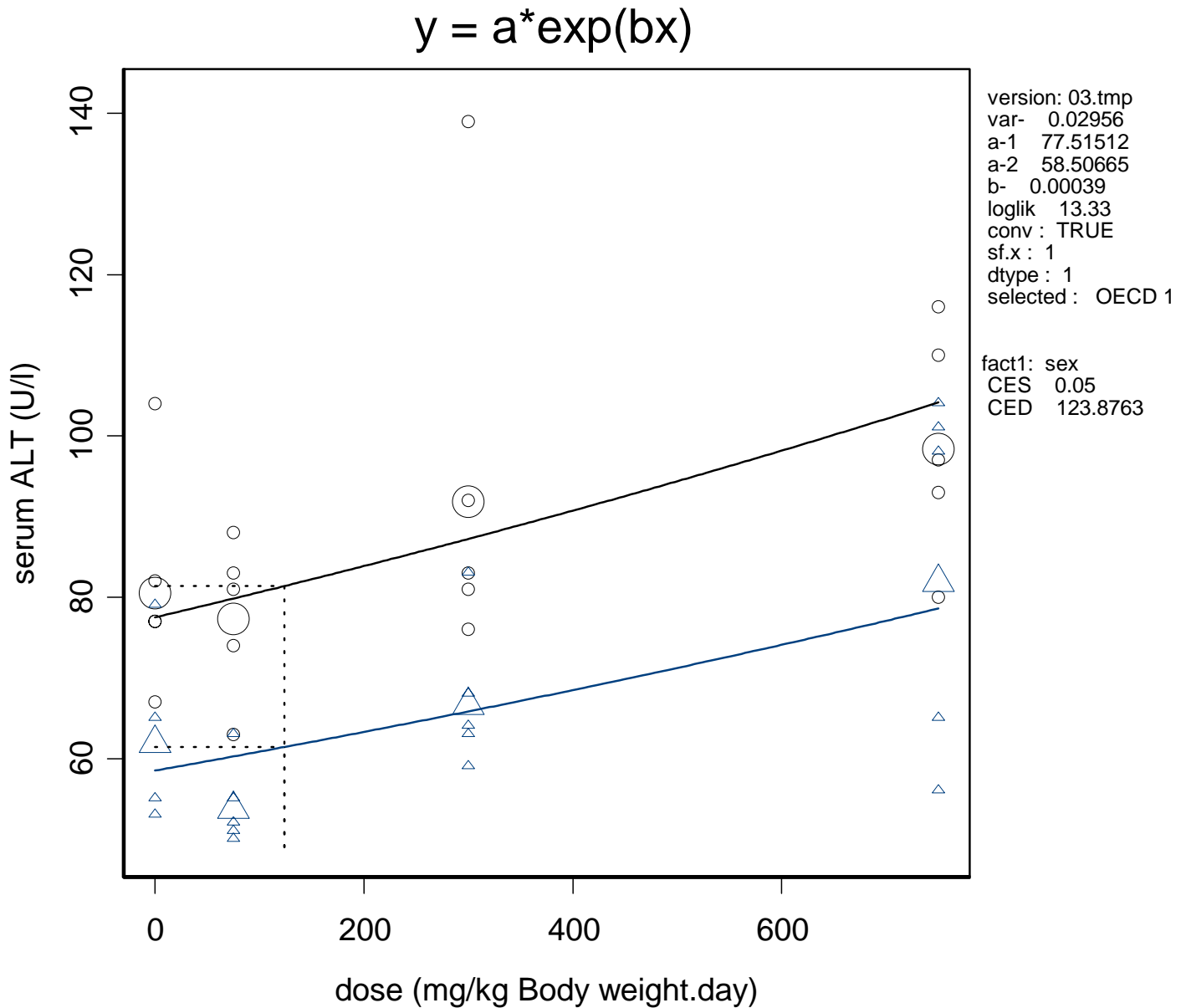
36 A threshold in the first sense may be incorporated into a model. The introduction of a threshold  
37 parameter truncates the dose-response relation at a threshold dose:

- 38  
39
- below threshold, the effective dose is zero
  - above threshold, the effective dose is the dose minus threshold
- 40

1  
2 Threshold terms generally do not improve model fit and confidence limits on thresholds are  
3 usually very large.  
4

#### 5 **5.2.1.4 Modelling with Covariates**

6  
7 In some circumstances, it is desirable to include variables in addition to an exposure variable in  
8 dose-response models. For example, in epidemiological studies, it is common to model disease  
9 risk in terms of not only exposure, but also age, gender, socioeconomic status, smoking status,  
10 and other measurements that may be relevant to the disease state. These other factors may not  
11 themselves be directly affected by the exposure, but they may be correlated with exposure status  
12 because of the way the sample was taken. Then, unless the proper covariates are included in a  
13 model for the relationship between exposure and the health endpoint, the effect of exposure will  
14 be incorrectly estimated. In bioassay studies, in which animals are randomized to treatment  
15 groups, this sort of confounding can not, in principle, occur, but it may be useful to include a  
16 covariate such as sex or body weight, to account for some of the variability in a related measure  
17 (see Figure 5.1).



1 **Figure 5.1 Dose-response model fitted to serum ALT levels observed in males (circles) and**  
2 **males (triangles), where sex is treated as a covariate. In this case, the**  
3 **parameter a (background response level) differs between sexes, while parameter**  
4 **b and the residual variance (var) for the (log-)data are assumed to be the same.**  
5  
6  
7

**Table 5.1 Continuous Dose-Response Models**

Name(s)	Notes	Equation for Response	Parameter Explanations
Michaelis-Menten Law of mass action	A theoretical account of enzyme or receptor based activity where the rate of action is a function of the rate of association (ka) and the rate of dissociation (kd).	$= R_{Max} \frac{[S]}{K_M + [S]}$	Rmax is the maximum rate of the reaction [S] is the substrate concentration KM is the Michaelis constant that is equal to ka/kd.
Hill equation Log-logistic	A modification of the Michaelis-Menten equation that supposes that the occupation of multiple sites or receptors is required for the production of an effect.	$= R_{Max} \frac{[D]^n}{K_D^n + [D]^n}$	Rmax is the maximum response [D] is the dose KD is the reaction constant for the drug-receptor interaction n is the power parameter
First-order Exponential	If the interaction between a chemical and a target site is irreversible, then the rate of the reaction is determined by the rate of association (ka) only.	$= R_{Max} (1 - e^{-rD})$	Rmax is the maximum response D is the dose r is the exponential rate constant
Power		$= \beta D^\alpha$	D is the dose $\alpha$ is the shape parameter $\beta$ is the scale parameter
Linear	Although there is usually no biological theory to suggest it, linear models are often justified by their simplicity – linear models have but a single parameter	$= Dm$	D is the dose m is the slope

3  
4  
5  
6

1  
2  
3**Table 5.2 Quantal Dose-Response Models**

Name(s)	Theoretical Basis	Equation for Frequency	Parameter Explanations
Step function	No variability	If $D < T$ , $f = 0$ If $D \geq T$ , $f = 1$	D is the dose T is the Threshold parameter
Single-hit Exponential	Hit theory models employ the use of a rate to describe the interaction between a group of causal agents (e.g. molecules) and a group of targets (e.g., a human population)	$= 1 - e^{-Dr}$	$R_{\max}$ is the maximum response D is the dose $K_D$ is the reaction constant for the drug-receptor interaction n is the power parameter
Multi-hit Gamma	An expansion of the exponential function, which is based on the notion that multiple hits or events are required to produce a particular effect.	$= \Gamma(\text{gamma} * D, k)$	$\Gamma()$ is the incomplete gamma CDF D is the dose gamma is a rate parameter k is the number of hits required to produce the effect.
Probit Normal	A descriptive model based on a normal or Gaussian distribution	$= \Phi(\alpha + D * \beta)$	$\Phi()$ is the normal CDF D is the dose $\alpha$ is a location parameter $\beta$ is the inverse scale parameter
Logistic	The statistical logistic model is also a descriptive tool with no theoretical basis.	$= \frac{1}{1 + e^{-\alpha - D * \beta}}$	D is the dose $\alpha$ is a location parameter $\beta$ is the inverse scale parameter
Weibull	A flexible descriptive model	$= e^{-(\alpha + (\beta * D)^\gamma)}$	D is the dose $\alpha$ is the background parameter $\beta$ is the inverse scale parameter

4  
5  
6  
7  
8**5.2.1.5 Biologically-Based DR Models**

9 While biological considerations may motivate the choice of one or several empirical  
10 models, the level of biological detail in such models is minimal. Thus, their credibility for  
11 interpolating and extrapolating a data set derives mainly from their fit to the data, as  
12 evaluated statistically. Another class of model, the biologically based dose response  
13 models (BBDR), are much more complicated, and are explicitly designed to model the  
14 biological details that lead from initial exposure to a toxicant to the ultimate pathological  
15 outcome. Typically, such a model includes a physiologically-based toxicokinetic model  
16 to describe the distribution and metabolism of the parent compound and toxic metabolites,  
17 and other mechanistic, or toxicodynamic, models that link target tissue concentration to  
18 the ultimate response. The toxicodynamic part of the models may be relatively simple,

1 e.g., when the outcome is inhibition of acetylcholine esterase in the model for  
2 chlorpyrifos (Timchalk, 2002), or may be as complicated as a fully elaborated stochastic  
3 model for carcinogenesis. Such a model is really a quantitative expression of a set of  
4 biological hypotheses, and when rigorously tested against critical experiments, becomes a  
5 credible tool for extrapolating from experimental results into exposure realms that are  
6 difficult or expensive to reproduce in controlled experiments. Such models are quite  
7 expensive both in resources and time to construct, and thus would be expected to be  
8 developed fully only for exposures and toxicities of the highest concern.  
9

## 10 **5.2.2 Statistical Distributions**

### 11 **5.2.2.1 Continuous Distributions**

12  
13 The normal or Gaussian distribution is symmetric and is defined from minus to plus  
14 infinity. It has two parameters – the mean and standard deviation, which control the  
15 location and scale of the distribution, respectively. Because sums of large numbers of  
16 small effects tend to be approximately normally distributed, this distribution is often used  
17 to describe variability and the variation of measurement error.

18 The lognormal distribution has two parameters, the geometric mean and the geometric  
19 standard deviation. It can be considered as a derivative of the normal distribution where  
20 the logarithms of the observed or predicted values are assumed to be normally distributed.  
21 This produces a skewed distribution on the original scale. Another consequence of using  
22 a lognormal distribution is that it will not generate negative values, which makes it more  
23 suitable for describing positive-only data sets and unsuited for values with negative  
24 values. Since many distributions are skewed and contain only positive numbers, the  
25 lognormal distribution often provides a good description. In addition, products of a large  
26 number of small effects tend to be approximately log normally distributed. Since effects  
27 in biological measures tend to be multiplicative (proportional) rather than additive, the  
28 lognormal distribution is generally more suitable for biological measures.

29  
30 The Weibull distribution is most commonly used to represent the survival or ‘lifetime’  
31 distribution of physical systems/products or biological systems depending upon the  
32 context. In many applications there is no explicit theoretical reasoning indicating that a  
33 Weibull distribution is appropriate or should be used, although the distribution does have  
34 some theoretical underpinning within the class of extreme value distributions. From a  
35 curve-fitting standpoint, the functional form of the distribution is simply a power  
36 transformation of the exponential model which gives the model more flexibility for  
37 describing data

38  
39 A more complete list of continuous distributions is given in Evans et al. (1993).  
40

### 5.2.2.2 Discrete Distributions

Discrete distributions describe the rates of occurrence for a finite set of possible outcomes. For example, they may be used to describe the behaviour of a tossed coin or a rolled die.

A Bernoulli distribution has outcome of 1 or 0, corresponding to the occurrence or absence of an event which occurs with frequency  $f$  over an infinite sequence of trials. The Bernoulli distribution is then simply '1' with frequency  $f$  and '0' with frequency  $1-f$ . The Bernoulli trial is the basis of the binomial distribution, the definition of which subsumes the former.

The binomial distribution is defined as the distribution of a sum of a given number of Bernoulli trials with outcome of 1 or 0, denoting the occurrence or absence of a specified event, respectively. In toxicological applications, the number of trials is fixed by the experimental design and the proportion of responding subjects occurring is the response to be estimated. As a result, the binomial distribution is the distribution typically used to estimate quantal response model parameters.

The Poisson distribution is a one parameter distribution for a positive and discrete valued response. The domain of the response variable is any positive integer. The distribution was originally derived as a distribution of rare events; specifically the number ( $n$ ) of events occurring in a sequence of Bernoulli trials where the number of trials is large and the probability ( $p$ ) of event per trial is small. Consequently, the Poisson can be used as an approximation of the binomial distribution when  $n$  is large and  $p$  is small. The Poisson distribution is commonly used in analyses of epidemiological data when the study design involves prospectively following a cohort of subjects over a time period for which the expected incidents of adverse events is small relative to the cohort size.

A more complete list of discrete distributions can be found in Evans et al., 1993.

## 5.3 Model Fitting and Estimation of Parameters

The general principles of parameter estimation and model fitting have been discussed in Chapter 3. Two basic methodologies are available for model fitting: conventional, in which parameters are selected to minimize or maximize an objective function, and Bayesian, in which information in a data set is combined with prior information about model parameters, resulting in a posterior distribution for those parameters that reflects the degree of uncertainty about those parameters. For historical and computational reasons, "user-friendly" software designed for carrying out dose-response analysis and nonlinear modelling in general has been restricted to using conventional methodologies, while Bayesian methods are implemented in packages that require more extensive programming and substantially greater understanding of the statistical details (for further details on Bayesian approaches, see Hasselblad and Jarabek, 1995; Gelman et al, 2004). While current software requires substantial statistical understanding for successful use of Bayesian methods, and are thus beyond the reach of this document, even conventional

1 methods require an understanding of some basic principles before outcomes from  
2 applying the software can be properly interpreted. Some general remarks may be helpful  
3 here.

### 4 **5.3.1 Criterion Function**

5 The general approach of fitting a model is to find parameter values for the model that  
6 optimize the fit of the model to the data. To that end, a criterion function is defined,  
7 reflecting the fit of the model. The goal is to find the parameter values that optimize the  
8 value of the criterion. For many models typically used, this can only be achieved by an  
9 iterative "trial and error" approach (see below).

10 In many applications, the logarithm of the likelihood function is used as the criterion.  
11 The likelihood directly derives from the distribution assumed for the scatter in the data.  
12 For quantal data, the binomial likelihood is typically used. For continuous data, the  
13 normal likelihood is often used, be it for the observed responses themselves, or for the  
14 log-transformed responses. Note that maximizing the normal likelihood function is in  
15 fact equivalent to minimizing the sum of squares.

### 16 **5.3.2 Search Algorithms**

17 Computer software use algorithms to find parameter values that optimize the fit of the  
18 model to the data, and the user does not need to worry about the exact nature of the  
19 calculations. However, some basic understanding of the search process is required in  
20 order to interpret the outcomes.

21 An iterative search algorithm tries to find "better" parameter values in a process by  
22 evaluating if the fit can be improved by changing the parameter values though a trial and  
23 error process. More advanced algorithms operate by evaluating the slope at which the fit  
24 is improved for one or more parameter value changes. The algorithm can only start  
25 searching when the parameters have values to start with. Although the software often  
26 gives a reasonable first guess for the starting values, the user may have to change these.  
27 It is not unusual (in particular when the information in the data is hardly sufficient to  
28 estimate the intended parameters) that the end result depends on the starting values  
29 chosen, and the user should be aware of that.

30 The algorithm keeps on varying the parameter values until criteria are satisfied for  
31 stopping. There are two major reasons for the algorithm to stop the searching process:

- 32 • The algorithm has converged (e.g., it has found a clear maximum in the log-  
33 likelihood function). In this case the associated parameter values can be  
34 considered as the "best" estimates (MLEs if the likelihood was maximized).  
35 However, it can happen that the log-likelihood function has not one but more  
36 (local) maxima. This means that one may get other results when running the  
37 algorithm again, but with other start values. This can be understood by  
38 remembering that the algorithm can only "feel" the slope locally, so that it  
39 usually finds the optimum that is closest to the starting point.

- The algorithm has not converged (i.e., the algorithm was not able to find a clear optimum in the likelihood function, but it stops because the maximum number of iterations (trials) is exceeded). This may occur when the starting values were poorly chosen, such that the associated model would be too far away from the data. Another reason could be that the information in the data is poor relative to the number of parameters to be estimated. For example, a DR model with five unknown parameters cannot be estimated with a four-dose-group study. As another example, the variation between the observations within dose groups may be large compared to the overall change in the DR. In these cases the likelihood function may be very flat, and the algorithm cannot find a point where the function changes between increasing and decreasing. The user may recognize such situations by high correlations between parameter estimates, i.e., changing the value of one parameter may be compensated by another, leaving the model prediction practically unchanged.

#### 5.4 Model Comparison

The fundamental criterion for judging a model is that the selected model should describe the data, especially in regions of the DR where inferences are needed. Most fitting methods provide a global goodness-of-fit measure, usually providing a p-value. These measures quantify the degree to which the model predictions correspond to the data. Small p-values indicate a poor fit to the data. Since it is particularly important that the data be adequately described, it is recommended that  $\alpha = 0.1$  be used to compute the critical value for goodness of fit, instead of the more conventional values of 0.05 or 0.01.

Another way to detect the form of these deviations from fit is with graphical displays. Plots should always supplement goodness-of-fit testing. For continuous data, it is extremely helpful that plots that include data points also include a measure of dispersion of those data points. In certain cases, the typical models for used in DRM cannot fit the observed data as, for example, when the data are not 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.

When fitting many different models to the same data, they generally will not all result in the same fit, and some care must be taken in choosing which model or models will be considered. In applying a statistical theory to this problem, one of four possible situations may arise:

1. The models form a nested series of models in the same family, in the sense that there is a “full” model, and other, “restricted” models are derived from that full model by setting successively more parameters to a fixed value, or conversely, successively incorporate more parameters into the model. Likelihood ratio tests can be used to evaluate whether the improvement in fit afforded by estimating additional parameters is justified. The general form of the test is to calculate  $2 \times (LL_{\text{full}} - LL_{\text{restricted}})$ , and compare this to a critical value from the chi-squared

1 distribution with  $P_{\text{full}} - P_{\text{restricted}}$  degrees of freedom (where  $P_x$  is the number of  
 2 parameters estimated in model  $x$ ).

- 3
- 4 2. The models are from the same family, but do not form a nested series. Some  
 5 statistics, notably Akaike's Information Criterion (AIC is  $-2L + 2p$ , where  $L$  is  
 6 the log-likelihood at the maximum likelihood estimates for the parameters, and  
 7  $p$  is the number of model degrees of freedom) can be used to compare models  
 8 (Akaike, 1973 ; Burnham and Anderson, 2002);. In this case, the model with the  
 9 smallest AIC value is selected, though models with similar AIC values (differing  
 10 by no more than around 4) are probably equivalent (Burnham and Anderson,  
 11 2002).
- 12
- 13 3. The models are not from the same family, but are fit using the same assumptions  
 14 about the underlying probability distributions, for example, all using a log-  
 15 normal likelihood, or all using a normal likelihood. In this case, Burnham and  
 16 Anderson (2002) argue that AIC can still be used to identify the best model, but  
 17 this appears to be a controversial point. In general, this case is still the subject of  
 18 statistical research. At present, it will probably be adequate to use AIC to select  
 19 a model as in the previous case, recognizing that this guidance may change.
- 20
- 21 4. Models do not use the same probability distribution. In this case, little formal  
 22 statistical guidance is available. The plausibility of assumptions about the  
 23 distribution of data needs to be examined by looking at the distribution of  
 24 individual data. However, often continuous data are aggregated and reported as  
 25 means and standard deviations, which eliminates the possibility of examining  
 26 distribution assumptions. In these situations, the best that can be done is to rely  
 27 on past experience with the endpoints being modelled, and select a reasonable  
 28 probability distribution accordingly.

29

30 As an example, consider a DR study that included the labelling index in the kidney of  
 31 female F344 rats exposed via gavage. The data are given in Table 5.3

32  
 33  
 34  
 35 **Table 5.3 Example dose-response study**

36

Dose mg/kg bw/day	Number of animals	Mean labelling index	Standard deviation, SD	Coefficient of variation, CV*
0	10	1.3	1.0	0.77
34	10	1.5	0.3	0.22
100	10	22.4	20.9	0.93
200	10	33.8	20.9	0.62

37 \*SD divided by mean

38  
 39 This is a continuous response, in the sense that the response can take on a (positive)  
 40 continuous range. Note that there is a large jump in the response between the lowest non-

1 control dose (34 mg/kg/day) and the next higher dose (100 mg/kg/day), and that the 34  
 2 mg/kg/day dose group deviates in showing much less variation than the other dose groups.  
 3 The response appears to level off (Figure 5.2), so that the dose-response curve fit to the  
 4 data needs to be sigmoid. The SD clearly increases with dose, but, except for the 34  
 5 mg/kg/day dose group, the CVs appear reasonably similar. This suggests using models in  
 6 which the coefficient of variation is taken to be constant across dose-groups.

7  
 8 Table 5.4 compares the fitting three models (including the Hill model) from PROAST  
 9 (Slob, 2004), run in the statistical program R (R Development Core Team, 2004), and the  
 10 Hill model from BMDS (USEPA 2001). PROAST assumes that the variation among  
 11 animals within a dose group can be approximated by a log-normal distribution, and that  
 12 coefficients of variation are the same in all dose groups, while BMDS assumes the  
 13 variation among animals within a dose group is approximated by a normal distribution,  
 14 the SDs being the same among dose groups. As long as the CV is not too large (e.g. less  
 15 than 10% - 15%) the lognormal is very similar to the normal distribution, and assuming  
 16 either distribution will give similar results. But at values for the CV similar to those in  
 17 these data, averaging about 65%, the two models for variation can give substantially  
 18 different results. Statistics from the two corresponding saturated models (calculating a  
 19 separate mean for each dose group, but holding coefficient of variation constant) is  
 20 shown for comparison.

21  
 22 **Table 5.4 Model Fitting**

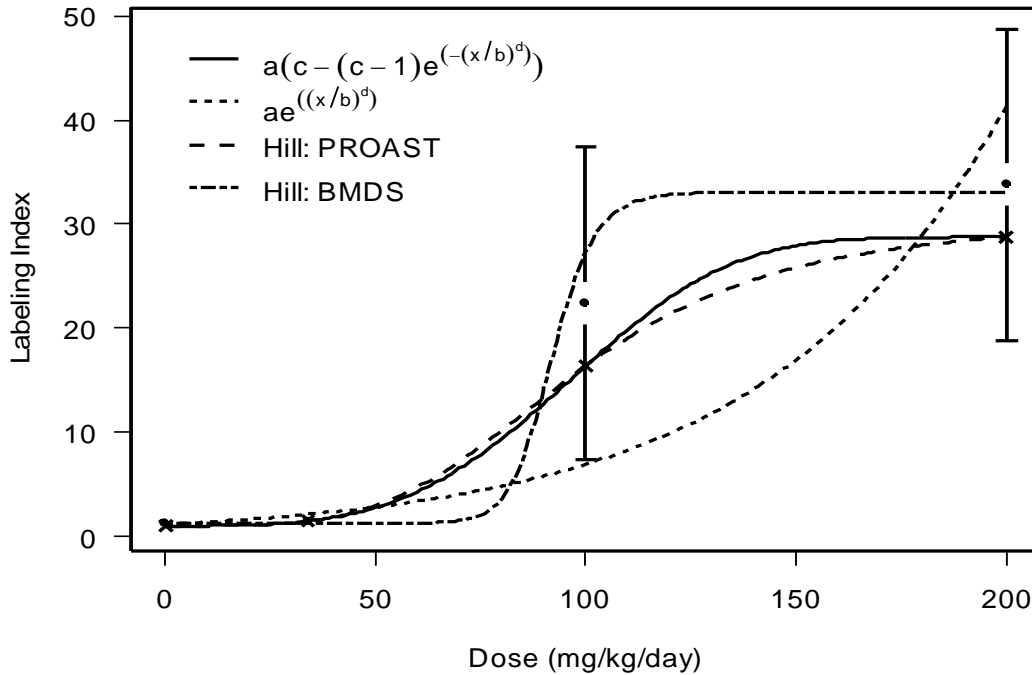
23

Model	Model df	LL	AIC	BMD	BMDL
PROAST Models					
Saturated	5	2.346	5.31	-	-
$a \cdot \exp((x/b)^d)$	4	-9.213	26.43	32.89	29.78
$a \cdot [c - (c - 1) \cdot \exp(-(x/b)^d)]$	5	2.346	5.31	40.48	29.56
Hill	5	2.346	5.31	39.96	NA <sup>1</sup>
BMDS Models					
Saturated	5	-72.58	155.16	-	-
Hill	5	-74.001	158.00	76.22	37.02

24  
 25 <sup>1</sup> NA = not applicable. The R version of PROAST was unable to calculate a  
 26 BMDL for this model. The BMDL from the S+ version of PROAST, based on  
 27 250 bootstrap samples, is 29.73

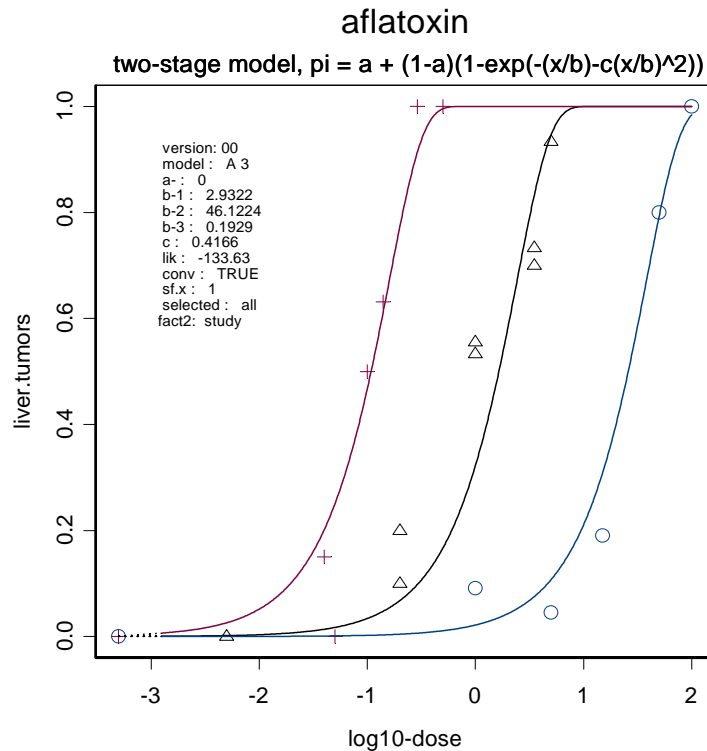
28  
 29 It is not possible on the basis of these statistics to directly compare the BMDS to the  
 30 PROAST models, because they are based on different probability models for the variance  
 31 among animals. However, the models within each software package can be compared  
 32 back to their corresponding saturated models. For example, although it is not possible to  
 33 compare the fit of the Hill model in PROAST to the fit of the Hill model in BMDS, both  
 34 models come very close to the fit of the saturated models, so are adequate descriptors of  
 35 the data (see, however, section 5.5.3). Of the models considered here, only the  
 36 exponential model (first PROAST model after the saturated model in the table above),

1 should be rejected since the P-value from comparing the log-likelihoods of this and the  
 2 saturated model is  $1.5 \times 10^{-6}$ , and the AIC values differ by over 20.  
 3



4  
 5  
 6 **Figure 5.2 DR models for data from Larson et al., 1995. Labelling index in kidney**  
 7 **of female F344 rats exposed via gavage. Error bars indicate 95%**  
 8 **confidence limits for the mean labelling index. 'X' characters indicate**  
 9 **geometric means**

10  
 11 When other data sets for similar endpoints exist, it may be possible to compare the results  
 12 of DRM across studies. As an example, Fig. 5.3 shows the two-stage model  
 13 simultaneously fitted to three different studies measuring liver tumour incidence as a  
 14 function of applied dose of aflatoxin.



1 **Fig. 5.3 Two-stage model fitted simultaneously to liver tumour incidence from three**  
2 **different studies (indicated by different symbols)**  
3

## 4 **5.5 Representing Uncertainty**

5  
6 Any parameters or predictions estimated from a given model are only point estimates and  
7 to a larger or smaller extent, uncertain. This uncertainty arises from at least three sources:  
8 (i) the sampling error arising from inferences about a larger population from a single  
9 experiment; (ii) the reality that DR estimates often differ among experiments with  
10 different experimental design, protocol, or uncontrolled circumstances, and (iii) the fact  
11 that the “true” model is not known, which results in additional uncertainty when  
12 interpolating between doses, but even more so when extrapolating outside the dose range  
13 containing observations. These three sources of uncertainty will be briefly discussed.

### 14 **5.5.1 Sampling Error**

15  
16 Uncertainty arising from sampling error with a single experiment is perhaps the easiest to  
17 evaluate and report. It may typically be quantified by a standard error or, preferably by a  
18 confidence interval. Confidence intervals may be calculated in several ways:

- 1
- 2 • plus or minus twice the parameter's standard error (provided by most dose-
- 3 response software), which is estimated by the second derivative of the likelihood
- 4 function (Hessian or information matrix);
- 5 • based on the profile of the log-likelihood function, using the Chi-square
- 6 approximation of the log-likelihood;
- 7 • bootstrap methods (see e.g., Efron 1987, Efron and Tibshirani, 1993); and
- 8 • Bayesian methods, in particular if one has some preliminary knowledge on the
- 9 plausible range of the parameter.

10  
11 Various studies have compared the first three methods and concluded that the first may  
12 result in inaccurate intervals, while the second and third methods give similar results (see  
13 e.g., Moerbeek et al., 2004).

### 14 **5.5.2 Study Errors**

15 Uncertainty about the true value of a parameter that stems from variability among  
16 experiments can often be handled by treating the experiments as comprising an additional  
17 level of hierarchy, when the experiments are very similar in design and intent (e.g. same  
18 agent on the same endpoint in the same strain and species; see Fig 5.3). To characterize  
19 uncertainty in a statistical framework, it can be assumed that there is a population of  
20 experiments, from which the ones at hand were selected (e.g., Davidian and Giltinan,  
21 1995). As a result, the prediction or parameter of interest varies around a mean value  
22 among the members of that population of experiments, and an estimate of the mean and  
23 the degree of confidence can be derived. It should be noted that, even if data from only  
24 one experiment is available for analysis, this source of uncertainty still exists – it may be  
25 possible to quantify this uncertainty by analogy.

### 26 **5.5.3 Model Errors**

27 The third area of uncertainty, model uncertainty, is reflected by the question: to what  
28 extent do the data, possibly along with other knowledge about DR shape, constrain the set  
29 of possible DR shapes? A statistical model completely hinges on the dose-response data,  
30 and the quality of the data is in fact the crucial aspect. In the fitting process a model tries  
31 to hit the response at the observed doses. However, when a model is used to make  
32 inferences, interpolation between observed doses, or even extrapolation beyond the non-  
33 control doses is considered. Thus, the model must also predict the response in the non-  
34 observed dose range. In other words, there are two aspects in evaluating the fitted model:  
35 one should not only assess if the model succeeded in describing the observed responses,  
36 but also if the model can be trusted to describe the non-observed responses where it is  
37 desirable to make inferences. The former aspect focuses on the quality of the model, the  
38 latter on the quality of the data. The following discussion elaborates on how to deal with  
39 the second of these two aspects (the first was addressed in section 5.4 – comparison of  
40 model fits).

41  
42 To evaluate if the data provide sufficient information for constraining the model may be  
43 done in two ways. The fitted DR model should be visually inspected, to check if the data

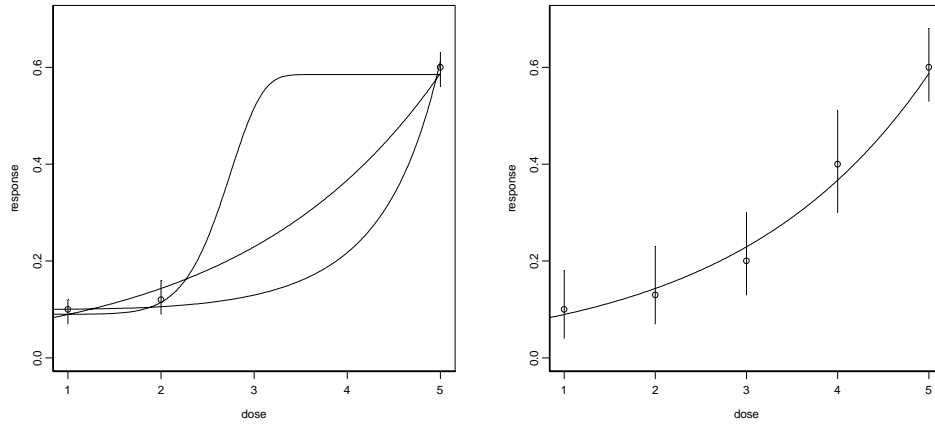
1 provide sufficient information to confine the model. Here, the question should be asked:  
2 if a curve is drawn through the data points by hand, could that be done in disparate ways?  
3 For instance, in the left panel of Fig. 5.4 three curves have been drawn through the data  
4 points, each of which might be close to the true dose-response curve. In the right panel,  
5 however, it is very hard to imagine that the true dose-response relationship would be very  
6 different from the (single) curve drawn here.

7  
8 Another way to deal with this question is by comparing the outcomes from different fitted  
9 models. If the data do contain sufficient information to confine the shape of the dose-  
10 response relationship, different models fitting the data (nearly) equally well, will result in  
11 similar fits and similar inferences. As an illustration, Fig. 5.5 shows two different  
12 models fitted to the same (continuous) data. Due to the good quality of the data, they  
13 result in very similar estimated DR relationships. Inferences from dose-response models  
14 bear an additional level of uncertainty in proportion to the degree with which those  
15 inferences depend on the model used.

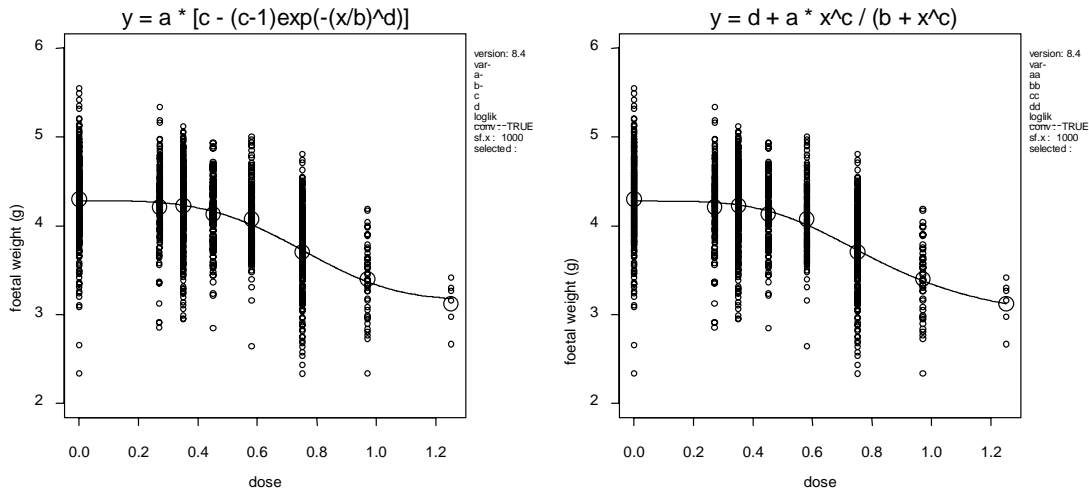
16  
17 In current practice, there is a tendency to focus only on goodness-of-fit, and passing a  
18 formal goodness-of-fit test is often regarded as sufficient evidence that the model is  
19 acceptable. This is unfortunate, since a goodness-of-fit test tends to be more easily passed  
20 for data with few dose groups, and exactly in that situation the second condition is more  
21 likely not to be met. In addition, a goodness-of-fit test assumes that the experiment was  
22 carried out perfectly, i.e., perfectly random with respect to all potentially relevant  
23 experimental factors and actions. Clearly, this assumption is not realistic.

24  
25 It is re-emphasized that a DR model, as long as it is not based on the mechanism of action  
26 of the particular chemical, only serves to smooth the observed dose-response relationship,  
27 and to provide for a tool to assess confidence intervals. A statistical regression model  
28 itself has little, if any, biological meaning, and the choice of the model is to some extent  
29 arbitrary. It is the data, much more than the model, that should determine the dose-  
30 response relationship, and any inferences derived from it. When different models (with  
31 similar goodness-of-fit and equal number of parameters) result in different estimates, this  
32 reflects a component of uncertainty that needs to be quantified and communicated with  
33 the estimate.

34  
35



**Figure 5.4** Two data sets illustrating the idea of model uncertainty. In the left panel the data (either quantal or continuous) do not contain sufficient information to confine the dose-response relationship: one may imagine various disparate curves that are all in agreement with the data, and hence they all might represent the true dose-response relationship. In the right panel the data points prohibit the possibility to draw disparate curves. Note that doses are indicated by group number.

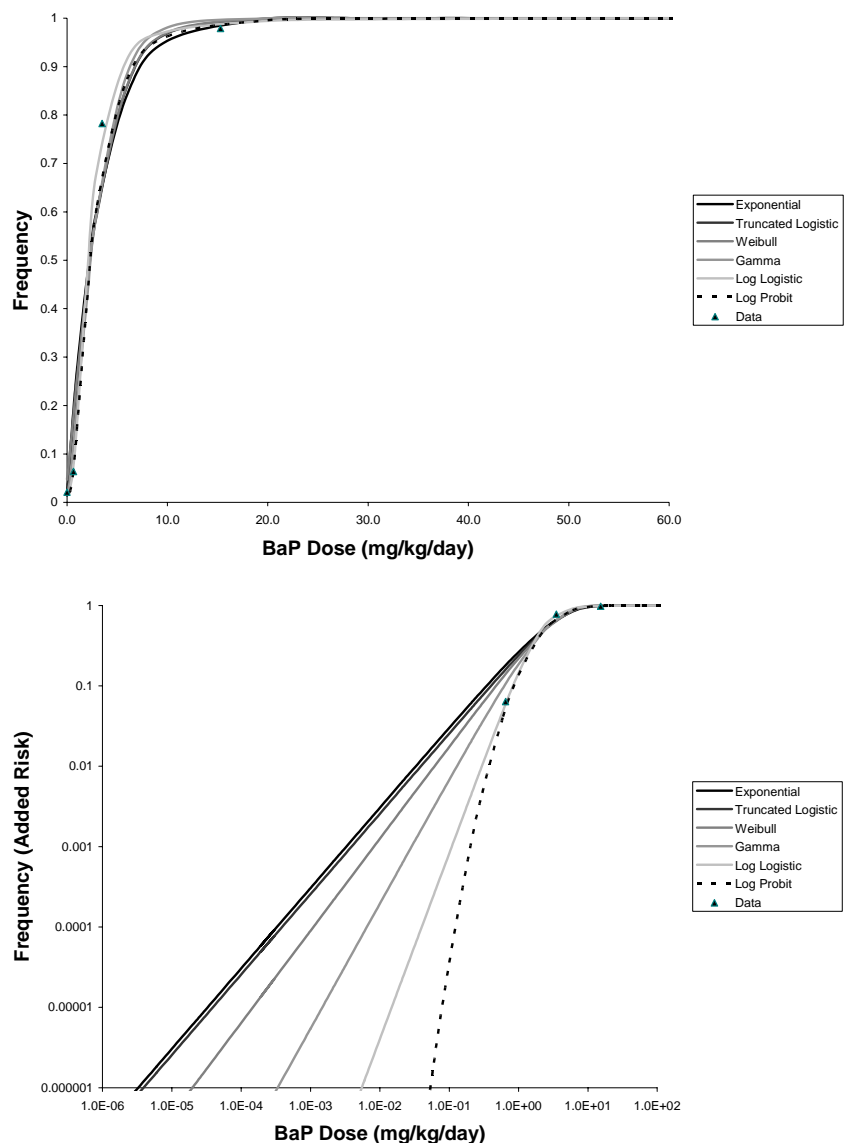


**Figure 5.5** Two different models (both with four parameters) fitted to the same data set resulting in similar dose-response relationships. Small circles indicate individual observations, large circles (geometric) means.

DR models that are based on the mechanism of action of the particular chemical stand in opposition to statistical models as described here. Such mechanistic models contain information gleaned from biological theory and typically multiple experiments, and therefore are less sensitive to data gaps (between dose groups). However, they do contain unknown parameters that need to be estimated from the data, and thus require the resulting uncertainties to be quantified. Since such models are typically complex and idiosyncratic, little further general advice can be given, and it is advised that professional statistical advice be sought in such cases.

Model uncertainty is particularly relevant to the issue of low-dose extrapolation. Here the problem is that there may well be several models that are consistent with the data, and so give similar predictions in the range of the data, but whose predictions diverge at the low end of the dose range (see Figure 5.6). One way to collect and represent model uncertainty in a risk assessment is through the use of probability trees (Rescher, 1967; Hacking, 1976). A probability tree is a logical construct that may be used to represent a set of mutually exclusive propositions. For example, if the three models depicted in the left panel of Figure 5.4 were equally well supported, then each model would have a probability of 1/3. If one model had a weight that was six times greater than the others, then it would have a probability of 0.75, while the others would have a probability of 0.125. Note that the probability of a model does not depend only on the strength of evidential support – it also depends on what other models are being considered. A model with little support may have a high probability if all the alternatives under consideration have even less support. Quantitative measures of model preference may be combined to produce an overall rank or to provide a formal measure of the weight-of-the-evidence.

To some extent, all quantitative methods for assigning model probabilities rely on untestable assumptions or elements of judgment. Therefore, the simplest and most straightforward method for assigning probabilities to models is to simply give them all the same weight. This approach is implicit when the predictions from different models are simply listed (e.g., Ghani et al., 2000). Another relatively simple approach is to ask the experts to identify plausible theories, and then apply probabilities to them (Evans et al., 1994; IPCS, 2000). These probabilities can then be updated to incorporate additional information in the data by using Bayesian methods. However, there are many formal techniques for assigning weights or probabilities to models (Bozdogan, 1987; Raftery et al., 1997). A semi-formal approach may be used in which the same criteria discussed in the section for selecting models, may also be used to weight and assign probabilities to each alternative model considered (e.g., Carrington and Bolger, 2000). Model uncertainty may also be integrated with sampling error by using bootstrapping techniques. This involves repeatedly drawing random samples from the data set and refitting each data set with a set of models. The best model(s) from each bootstrap are then retained in a probability tree to represent both parameter and model uncertainty.



**Figure 5.6. Model uncertainty in low dose extrapolation. Different models may all fit the data reasonably well (top), but yield highly divergent response estimates at low doses (bottom). The data and models are taken from Fitzgerald et al., 2004.**

## 5.6 Benchmark Dose and Benchmark Response Selection

One important use of DRM is the calculation of BMDs. A BMD is the dose at which it is inferred that a particular, prespecified level of response would occur. The methodology was introduced in Crump (1984) as an alternative to the use of NOAELs and LOAELs in dose-response assessment for determining quantities such as acceptable daily intakes (ADIs). The main advantages of the use of the BMD over NOAELs and LOAELs stems from the more complete use of information in the dose-response data by BMD methods, and from the fact that uncertainties about the value of a BMD can be quantified using statistical methodology.

The uncertainty of a BMD may be expressed as a confidence interval, in which case the lower end of a one-sided 95% confidence interval is termed the BMDL, or as a full Bayesian posterior distribution.

The Benchmark Response (BMR) is the response for which the BMD is to be calculated. There are both technical and policy aspects to selecting the BMR. The technical aspects have to do with just how the BMR is expressed – different types of endpoints, such as quantal and continuous, require different treatments. Also, in somewhat more complicated situations, such as when covariates have been used in the modelling, the BMD depends on the BMR and possibly on the values of the covariates. Policy issues have to do with just how high or low down the dose-response curve the BMR should be. This section discusses the technical issues surrounding the choice of BMR, and some of the consequences that need to be considered in making the policy decision about where to set the BMR, but does not directly address the choice of its particular value.

#### Technical Issues of Expressing the Benchmark Response

The way the BMR is expressed depends upon the kind of response variable being modelled. For endpoints with two states (affected – not affected), the BMR is usually expressed in a way that adjusts for background. Two equations are common. One is that of added risk:

$$BMR_{AR} = f(BMD) - f(0)$$

where  $f(x)$  represents the dose-response function evaluated at dose  $x$ . The other, which is probably most widely used, is extra risk:

$$BMR_{ER} = \frac{f(BMD) - f(0)}{1 - f(0)}$$

where added risk is divided by the non-affected fraction of the non-exposed population. The response at the  $BMD_{ER}$  is always smaller than the response at the  $BMD_{AR}$  for the same numerical value of BMR when there is a background incidence. However, for small to moderate background response the difference is small.

A third equation, common in epidemiological analyses, but applicable to animal studies as well, is relative risk:

$$BMR_{RR} = f(BMD) / f(0)$$

BMRs for continuous endpoints can be expressed directly in terms of changes in the mean response level, or indirectly in terms of the fraction of experimental animals that exceed (or drop below) some critical level. For example, the BMD for mean adult body weight might be selected to be the dose at which the mean body weight drops below 90% of the body weight in controls, or at which brain acetylcholine esterase activity is inhibited by 10% relative to control levels (this is often termed the Critical Effect Size, CES). One might also specify a fixed value or fixed drop in the mean, selecting, for example, the dose at which the mean nerve conduction velocity drops below a fixed rate, or a fixed difference from that in unexposed individuals. For endpoints that demonstrate a sigmoid response, as does enzyme induction, it has been suggested (Murrell et al. 1998; see Gaylor and Aylward, 2004, for a

contrary argument) that a formulation similar to extra risk be used: for these endpoints, the authors suggest that the BMD is best characterized as the dose at which the response is a specified fraction of the total dynamic range (the difference between background and maximum possible induction, for example) of the response.

Indirect or “hybrid” approaches have been advocated by Crump (Crump, 2002) and Gaylor and his coauthors (Gaylor and Slikker 1994; Kodell et al. 1995). In indirect approaches, the relationship between the mean of a continuous variable and dose is modelled, in the same manner as in the direct approaches. Next, a critical value for the continuous variable is determined which is to be considered as adverse, and an extra (or additional) risk BMR is selected for which to calculate a BMD. It is preferable that the critical value be based upon biological considerations, but, it may otherwise be a value in the tail of the distribution of values in the control group. As the mean response increases, so will the fraction of subjects that exceed the previously determined critical value. The BMD is the dose at which the fraction exceeding the critical value corresponds to the fraction of affected animals associated with the BMR as defined for quantal data (e.g.  $BMR_{ER}$ ).

It is possible to approximate the BMD as calculated in the previous paragraph (Crump 1995) for a critical value corresponding to a “small” (e.g. 0.1 – 2%) risk in the control group and extra risk in the vicinity of 10%. This BMD corresponds approximately to the dose at which the mean of the response variable differs from the control mean by an amount equal to the standard deviation of the control group. This gives another way to specify a BMR for continuous variables, based on the variability of the animals used in the bioassays.

Both hybrid methods based on variability discussed above require that the variability be true inter-individual variability, and not be due to large assay errors. They depend critically on the idea that extreme quantiles of an unexposed population may be thought of as affected in the same sense as an individual with the same value from an exposed population.

In some cases, the dose is not the only independent variable in a dose-response model. For example, in epidemiological studies, often many covariates that help characterize an individual and that might influence the response variable and be incidentally associated with the exposure variable are included in analyses in an attempt to reduce bias in the estimates of the effects of exposure (see section 5.2.1.4). In developmental bioassays, characteristics of the dam or the litter as a whole (such as number of implantation sites), may be used as a covariate in the modelling to help explain some of the additional variation among litters usually seen in such studies. Even adult-only rodent bioassays are usually segregated by gender. Typically, then, the assessor needs to decide for which values of the covariates BMDs need to be calculated. When there are few, discrete covariates, it may make sense to calculate a separate BMD for each set of values (e.g., a BMD for both males and females). When covariates are continuous (or treated as such, as in number of implantation sites), in an animal bioassay, it is usual to pick a typical value in the control group. However, if BMD changes with the value of the continuous variable, and the variable makes sense for extrapolation to the human situation, BMDs should be calculated for several reasonably extreme values of the covariate.

## 5.7 Summary

Data sets for DRM generally need to be selected to reflect the more sensitive endpoints available, just to reduce potential workload. Models used depend upon the type of data (continuous, ordered categorical, quantal, or counts), and include a model for DR and a model for the variability of the data. Once models are fit to a data set, the degree to which they individually describe the data is evaluated using goodness of fit measures, and in addition, their ability to describe the data with respect to each other may be compared using measures such as the Akaike information coefficient. Uncertainty about the inferences that result from such model fall into three main categories: statistical uncertainty of inferences due to variability among responses in experimental subjects, variability among experiments due to unavoidable differences in experimental execution, and uncertainty due to the fact that different models yield different approximations of the true DR relationship. DR analysis needs to address all three sources of uncertainty whenever possible. One particular important application of DRM is the calculation of benchmark doses, doses at which it is inferred that a particular level of response would occur. When data are available, benchmark doses are a better alternative to NOAELs or LOAELs in the calculation of guidance values such as ADIs or TDIs. When extrapolation is necessary, the uncertainty associated with any predictions made should be represented. It is often especially important to include model uncertainty.

## **6. COMMUNICATING THE RESULTS OF DRM**

### **6.1 Introduction**

Risk communication has been defined as the "interactive exchange of information about (health or environmental) risks among risk assessors, managers, news media, interested groups and the general public (WHO, 2004b). Risk communication has evolved with the rest of the risk analysis paradigm to embrace the "interactive" nature of the processes. The transition from monolog to reflexive dialog in risk communication has necessitated awareness that risk perception issues are extremely important. The scientific, political and social perspectives of bench scientists, risk assessors, risk managers, media and the public, respectively, can result in considerable misunderstandings and misinterpretations (Garvin, 2001). The preconception that scientific and technical knowledge and their application in risk analysis are value-free and objective has often resulted in the marginalisation of insights from other sources.

General public perception, resulting from health based guidance approaches and terminology such as "ADI", "TDI", and "threshold", is that there is a bright line between "safe" and "unsafe". These approaches are not designed to incorporate risk and benefit dynamics, and may not require or even allow an outside audience to become engaged in the decision process. For many considerations of chemical exposures, these dynamics do not have to be dealt with because the outcome of the safety/risk assessment provides a perfectly useful and acceptable answer to the risk manager. However, there are instances where these dynamics will need to be considered and evaluated.

The use of DRM and other probabilistic assessment techniques to quantitatively describe variability and uncertainty brings new challenges in risk communication. Some of these challenges are

- explaining that a certain percentage of the population is predicted to experience some effect,
- explaining the level of risk in those circumstances where there is no safe level of exposure,
- comparing competing risks or benefits,
- providing a focus on uncertainties that are attendant to the predicted risk, and
- explaining that the risk generally is described at the population level, rather than the individual, noting that this is also the case for the ADI/TDI approach.

In addition, one of the limitations of the current health based guidance approach is that it gives no information about risk when the ADI/TDI is exceeded. For example, some subpopulations may exceed the TDI for dioxins, and the DRM approach may provide additional information that is useful for the risk manager and communicator.

An appreciation of the variability in most populations clearly impacts risk communication. This is particularly true for genotoxic carcinogens and other substances, such as lead, that are unavoidable contaminants and may be toxic at extremely low levels. Using a point estimate to depict an entire population in the context of risk communication can be misleading, because it can suggest that the risks are larger for the entire population than they really are if upper percentile point estimates are used and it ignores the fact that some portion of the population does have a somewhat higher level of risk. Becoming involved in a public decision requires a transformation from concern for an individual to concern for a population,

and thinking about variability as an inherent part of the problem rather than just a source of uncertainty.

In risk communication, uncertainties can facilitate dialogue. Uncertainty analysis can inform all the parties of what is known, what is not known, and the weight of evidence for what is only partially understood. However, there are currently no general criteria for the application of weight of evidence approaches. An appreciation of uncertainty, including uncertainty about variability, can lead to better consideration of the options for seeking better information, using a value-of-information approach (Thompson, 2002). However, in risk communication, "uncertainty" can be a double-edged sword. When the results of a probabilistic risk assessment are presented, uncertainty is specifically described rather than managed by the use of a default factor. Since the responsibility for managing the uncertainty is left to the discretion of the management process, communicating the uncertainty to the participants in that process is very important.

The application of DRM and other probabilistic risk assessment techniques has the potential for improving risk analysis and public risk perception. There must be an acknowledgement of the limitation and weaknesses of the technical knowledge in addition to its strengths. There also should be the realization that there may be difficulties with risk comparisons, and that social perceptions can drive precautionary considerations. There may not be agreement on how to interpret new information or on the appropriate criteria for making or reversing risk decisions. The critical contribution of probabilistic approaches is that they can improve the transparency of the processes of risk assessment and risk management and thereby facilitate communication. As a result, participation in the decision process will be broadened.

### **6.1.1 Incorporation of the outputs of DRM into risk assessment**

The output of dose-response analysis can be used in various ways dependent on problem formulation and the nature of the effect modelled. An output may be presented in three principal ways as the basis for advice on the possible health implications of human exposure

- establishment of a health-based guidance value, such as an ADI or TDI that is a daily intake over lifetime that is considered to be without appreciable health risk (This would be analogous to current procedures based on a NOAEL or LOAEL.),
- estimation of the margin of exposure (MOE) as the ratio between the DR output and the estimate of human exposure, and
- quantitative estimation of the magnitude of the risk at the level of human exposure, derived from the modelled dose-response relationship.

The discussion below assumes that the dose used in the DR model was the external dose expressed in mg/kg body weight. The use of internal or target-organ dose estimated by a physiologically-based toxicokinetic (PB-TK) model would reduce the uncertainties of interspecies extrapolation because kinetics are a major source of species differences, such that a reduced uncertainty factor would be required.

### **6.1.2 Derivation of health-based guidance value**

Traditionally a health-based guidance value for threshold effects has been derived from a NOAEL or LOAEL divided by an appropriate composite uncertainty factor, either default values or chemical-specific adjustment factors (CSAFs), on the assumption that the NOAEL

represents an intake close to the threshold for the adverse effect. In practice, the limit of detection for the incidence of adverse effects in animal experiments is approximately 5%, and many studies have shown that the BMDL for 5% response is similar to the experimental NOAEL (Allen et al., 1994). Fowles et al. (1999) came to a somewhat different conclusion. They examined acute inhalation lethality data, and compared NOAELs to benchmark doses corresponding to 1%, 5%, and 10% response incidences. Similarly to the “quantal” parts of the results of the Allen et al. (1994) studies, BMDLs based on 10% incidence corresponded approximately to NOAELs. However, because the DR for lethality is so steep, BMDLs for 5% and 1% incidences were very close to those for 10% incidence. As a result, the BMDLs for a 1% incidence were on average only about 1.6 or 3.6 times smaller than a NOAEL, depending on whether a log-probit or Weibull model was used. Therefore, given the uncertainty in the relationship of the NOAEL and the threshold of the adverse effect, finding a BMR such that the resulting BMD and BMDL correspond numerically (on average) to a NOAEL may not be relevant and is certainly not necessary for the application of BMD approaches. Also, the use of the BMDL to set a health-based guidance value would need to take into account the same uncertainties as when a NOAEL is used as the basis for establishing an ADI/TDI.

### **6.1.3 Estimation of the MOE.**

The normal default uncertainty factor of 100 has a long history of use for threshold effects and can be regarded as the margin between two points – the NOAEL or BMDL from the experimental data, and a level of human intake/exposure that would be without appreciable health risk. Because this is based on a NOAEL or BMDL the ratio is equivalent to a margin of safety (MOS), and there would be negligible risk providing that the intake was at or less than the ADI/TDI.

In the case of adverse effects that are considered not to show a biological threshold in their dose-response, such as those induced by a genotoxic carcinogenic compound, the BMDL could not be considered to represent an intake close to a threshold, but is simply the confidence interval on the BMD. Consequently, the margin between the BMDL and the estimated human intake/exposure would not be a MOS and is therefore termed a margin of exposure (MOE). Calculation of the MOE is the ratio between two experimental estimates, the BMDL and the predicted or estimated human intake/exposure. Calculation of an MOE does not require extrapolation the data beyond the range of observations (WHO, 1999 and Edler et al., 2002)

Uncertainties related to interspecies differences and human variability which are the basis for the usual 100-fold uncertainty factor used in the derivation of an ADI/TDI would be equally applicable to an MOE based on animal data, but there would be additional uncertainties related to the nature of the dose-response relationship below the experimental/observable range, the impact of genetic polymorphisms in the processes critical to the production of a mutated cell, and the subsequent clonal expansion and progression into a cancer. Consequently, an MOE of 100 would be inadequate to reflect the fact that the starting point (the BMDL) cannot be regarded as a threshold or the additional uncertainties related to the mode of action.

Application of linear low-dose extrapolation using the BMDL for a 5% response (see below) to estimate a 1 in a million lifetime risk is equivalent to a MOE of 50,000.

#### **6.1.4 Quantitative estimations of the magnitude of the risk at levels of human exposure.**

The results of a DR model can be used to estimate the possible risks at intakes/exposures above a health-based guidance value such as the ADI and at very low levels of human exposure, or to estimate intakes/exposures associated with predefined levels of risk, such as a one in a million lifetime risk of cancer.

Estimation of risks of intakes/exposures above a health-based guidance value, derived by the application of uncertainty factors to a BMDL from a study in either animals or humans, would need to use the slope characteristics in the DR model. For example, if an intake is of concern because it is above the health-based guidance value, then the extent of any risk could be estimated by reference back to the modelled animal DR relationship. If the intake were twice the ADI/TDI, then the risk would be given by the response estimate predicted by the model at twice the BMDL. Traditionally, an estimate of the possible risk has not been made and intakes above the ADI/TDI have been considered to have eroded the uncertainty factor.

Estimation of risks at very low levels of human exposure, or of exposures associated with responses below the BMR, requires extrapolation outside the data used to generate the DR model. Extrapolation outside the observed range, for example from an incidence of about 5% to one in a million will require extrapolation over many orders of magnitude. Low-dose extrapolation may be undertaken using the DR relationship(s) defined by the model(s) that were fitted to the experimental data or by application of a standardized mathematical approach, such as linear extrapolation, to the starting point. An advantage of using a consistent model is that the risk estimates can be compared across different compounds. The major uncertainty associated with such estimates is the biological relevance of the model in the region of extrapolation.

### **6.2 Presentation of Results**

In a scientific or logical sense, the risk assessment is finished when the conclusions have been drawn. However, when the conclusions are simulation results, some distillation or condensation is often necessary in order to make the results comprehensible. Since there is always some danger that crucial information may be lost, care must be exercised to ensure that the summary process does not omit information that is important for the decision

#### **6.2.1 Tables**

Precise communication of quantitative information requires numbers. More numbers will portray more information than fewer numbers, but will take longer assimilate. Tables 6.1-6.5 give examples of the range of options that may be considered, all taken from the same simulation results. It is recommended that, in case of effects of concern or a single effect found in several studies, all quantitative results are summarized in a table.

**Table 6.1 Percentiles from a two-dimensional simulation**

		Uncertainty												
		Average	SD	Minimum	Perc 0.01	Perc 0.05	Perc 0.10	Perc 0.25	Median	Perc 0.75	Perc 0.90	Perc 0.95	Perc 0.99	Maximum
Variability	Average	0.457	0.063	0.234	0.236	0.366	0.403	0.456	0.462	0.497	0.502	0.503	0.510	0.874
	Minimum	0.047	0.061	0.000	0.000	0.000	0.000	0.016	0.055	0.076	0.076	0.076	0.076	0.874
	Perc 0.01	0.094	0.065	0.000	0.000	0.000	0.007	0.072	0.101	0.129	0.129	0.130	0.131	0.874
	Perc 0.05	0.146	0.068	0.000	0.000	0.000	0.069	0.144	0.148	0.178	0.179	0.180	0.180	0.874
	Perc 0.10	0.188	0.074	0.000	0.000	0.000	0.116	0.187	0.205	0.216	0.216	0.217	0.218	0.874
	Perc 0.25	0.274	0.083	0.000	0.000	0.119	0.207	0.287	0.291	0.317	0.320	0.320	0.327	0.874
	Median	0.401	0.105	0.000	0.000	0.267	0.352	0.399	0.404	0.471	0.476	0.476	0.484	0.874
	Perc 0.75	0.586	0.064	0.388	0.394	0.519	0.531	0.561	0.568	0.651	0.657	0.657	0.667	0.874
	Perc 0.90	0.808	0.030	0.760	0.762	0.774	0.776	0.784	0.790	0.843	0.847	0.848	0.858	0.874
	Perc 0.95	0.949	0.024	0.874	0.923	0.930	0.931	0.941	0.944	0.953	0.963	1.014	1.056	1.058
	Perc 0.99	1.247	0.086	0.874	1.138	1.142	1.147	1.149	1.287	1.296	1.321	1.403	1.462	1.473
	Maximum	2.192	0.483	0.875	1.573	1.579	1.584	1.599	2.559	2.592	2.608	2.619	2.663	2.670

**Table 6.2 Population percentiles with confidence intervals**

Average	0.457 (0.366, 0.503)
Minimum	0.047 (0.000, 0.076)
1 <sup>st</sup> Percentile	0.094 (0.000, 0.130)
5 <sup>th</sup> Percentile	0.146 (0.000, 0.180)
10 <sup>th</sup> Percentile	0.188 (0.000, 0.217)
25 <sup>th</sup> Percentile	0.274 (0.119, 0.320)
Median	0.401 (0.267, 0.476)
75 <sup>th</sup> Percentile	0.586 (0.519, 0.657)
90 <sup>th</sup> Percentile	0.808 (0.774, 0.848)
95 <sup>th</sup> Percentile	0.949 (0.930, 1.014)
99 <sup>th</sup> Percentile	1.247 (1.142, 1.403)
Maximum	2.192 (1.579, 2.619)

**Table 6.3 Population percentiles with standard deviations**

Average	0.457 ± 0.063
Minimum	0.047 ± 0.061
1 <sup>st</sup> Percentile	0.094 ± 0.065
5 <sup>th</sup> Percentile	0.146 ± 0.068
10 <sup>th</sup> Percentile	0.188 ± 0.074
25 <sup>th</sup> Percentile	0.274 ± 0.083
Median	0.401 ± 0.105
75 <sup>th</sup> Percentile	0.586 ± 0.064
90 <sup>th</sup> Percentile	0.808 ± 0.030
95 <sup>th</sup> Percentile	0.949 ± 0.024
99 <sup>th</sup> Percentile	1.247 ± 0.086
Maximum	2.192 ± 0.483

**Table 6.4 Selected population percentiles with confidence intervals**

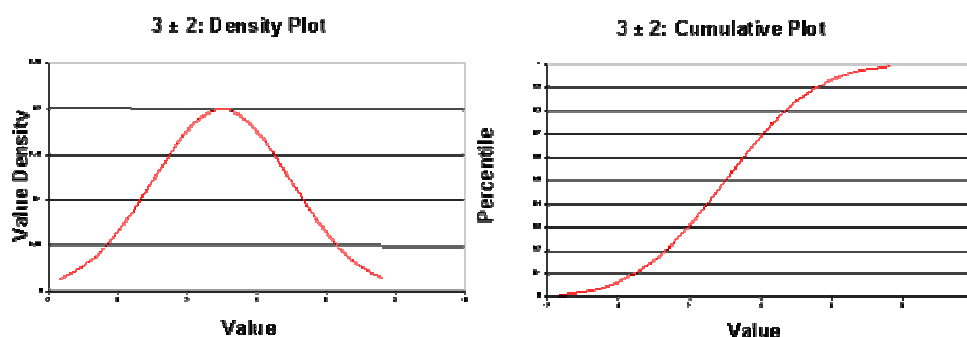
Average	0.457 (0.366, 0.503)
Median	0.401 (0.267, 0.476)
90 <sup>th</sup> Percentile	0.808 (0.774, 0.848)
95 <sup>th</sup> Percentile	0.949 (0.930, 1.014)
99 <sup>th</sup> Percentile	1.247 (1.142, 1.403)

**Table 6.2. Population mean with uncertainty estimate**

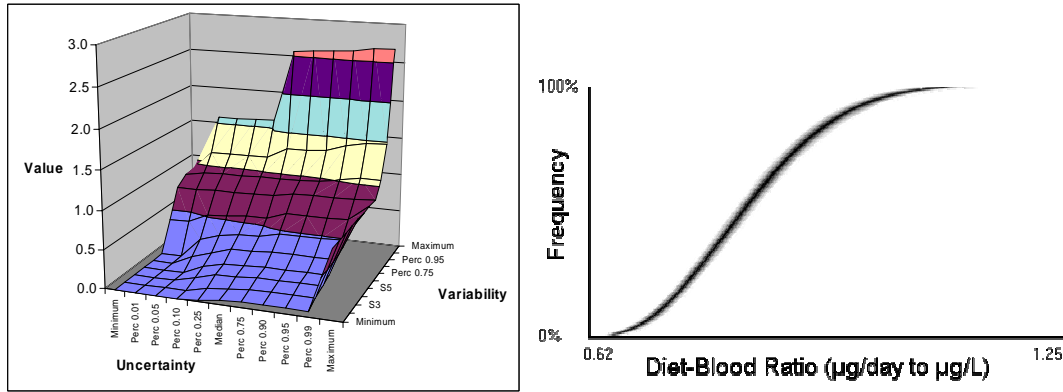
Average	$0.457 \pm 0.063$
---------	-------------------

### 6.2.2 Graphs

Although they may allow quick comparison, tables inherently compare one value at a time. Graphing or visualization is in some ways a better means of digesting the entire distribution. A one-dimensional simulation will produce a frequency distribution (when simulating variability) or a likelihood distribution (when representing uncertainty). There are two ways of plotting frequency or likelihood curves (see Figure 6.1). The first is to plot density vs. value, which emphasizes the values which are the most common or likely. The second is to plot cumulative percentiles vs. value, which allows the percentile corresponding to a particular value to be read from the plot. A graphical presentation of the dose modelling in relation to the experimental data may also be helpful in deciding on which dose-descriptor should be used for lifetime risk.

**Figure 6.1 Plotting frequency distributions**

Two-dimensional results are more difficult to display. Two strategies for adding an extra dimension are illustrated in Figure 6.2. The first uses three-dimensional perspective to portray the third dimension. The second uses shading, where darker hues are used to represent either higher density or more central values. This is particularly of use for displaying uncertainty, as the less well-defined (more uncertain) parts of a curve appear fuzzy.



**Figure 6.2** Plotting results of three-dimensional simulations

### 6.3 Risk Assessment Context and Questions

The output of the DRM should be directed towards addressing specific questions about the likelihood of adverse health effects in response to exposure to chemicals. This would build on conventional safety assessment procedures that have been accepted internationally as the indicator for determining safe levels of exposure. These rely on the identification of a NOAEL/NOEL for a critical endpoint in the effect data and incorporation of uncertainty factors to allow for inter-species and inter-individual variation.

DRM offers the potential to provide additional information for the risk manager, specifically a more scientifically robust method for determining the health based guidance values (e.g., ADI) using the BMD and better information on the likelihood of effects at low doses that are below the levels observed in biological systems. The mathematical models will also provide estimates of the statistical uncertainty surrounding estimates of likely effect.

Whether traditional safety based or DRM assessments are carried out, the risk manager will still require information on the toxicology of the adverse health effect and the robustness of the determination of the health based guidance value to help inform the management options. This may include the following:

- a discussion of the strength and weight of evidence
- uncertainties and gaps in the data
- information on the nature and severity of (critical) effect(s)
- limitations in the interpretation
- assumptions made in the analysis
- qualitative assessment of the potential effects of exceeding the health based guidance value

## 6.4 Synopsis of Approach to Modelling

Dose-response modelling involves six basic steps, i.e., data selection, model selection, statistical linkage, parameter estimation, implementation and evaluation (see Chapter 3, Table 3.1). In undertaking a DRM exercise, several of these factors that will impact the types of outputs and that may be of importance to the risk manager are briefly described below.

### 6.4.1 Data sets

Traditional safety assessments focus primarily on a single critical end point, whereas dose-response modelling gives the potential for separating out multiple end points. Modelling outcomes may be based on data from single or multiple experiments. In the latter situation, meta-analysis may integrate the results of several independent studies that are considered to be ‘combinable’.

The risk manager could see four types of data from the modelling evaluations including quantal, count, continuous and ordinal categorical data. The risk manager will need to understand what data sets were modelled and if quantitative information from more than one dataset is presented will need guidance on the rationale for forwarding the additional dataset information and for synthesizing this additional information. This guidance may include information about consistency (or inconsistency) of quantitative response across the endpoints. Such information could be used by the risk manager to strengthen (or weaken) their confidence in the quantitative evaluation of the potential for health impacts.

If DRM information is available from human epidemiology evaluation, then an understanding of both the strengths as well as possible limitations (often in the quantitative exposure information) in the dataset may also temper or strengthen the qualitative or quantitative assessment from the animal studies.

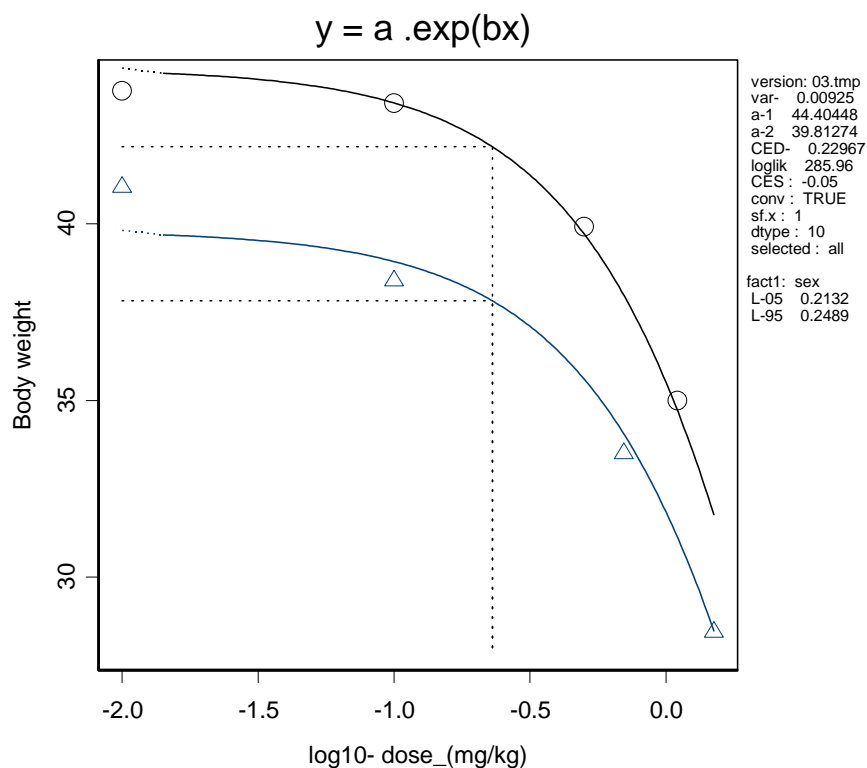
### 6.4.2. Severity (Degree of Effect)

The severity of toxic responses is rarely used in DRM other than in a qualitative manner, e.g. tumour formation vs. reduced fertility. However, one may also consider severity or degree of response in a quantitative way at the level of a single endpoint. As discussed in Chapter 5, the dose-response of continuous endpoints may be directly interpreted as a dose-related change in degree of effect, e.g., a percent decrease in haematocrit (Woutersen et al., 2001), or a percent change in body weight (Fig. 6.3, which represents the dose-response relationship of body weight (BW) for exposure to the mycotoxin deoxynivalenol (DON) (Pieters et al., 2004). Here, a certain degree of effect (5% reduction in BW) is chosen for deriving the BMD (or CED=critical effect dose). The BMDL then defined as the dose associated with a particular (e.g. 5%) change in degree of effect for that endpoint.

An important advantage of defining a BMR in terms of degree of effect based on continuous response data is that values for the BMR that may be considered non-adverse are within or close to the range of observations. Therefore, low-dose extrapolation may not be needed, or only to a small extent when continuous endpoints are considered.

In the case of a histopathological endpoint resulting in ordinal data, a dose-response function may be fitted using categorical regression, and the BMDL associated with a particular degree of effect (e.g. minimal, or mild) may be estimated (e.g.; Piersma et al., 2000; Woutersen et al., 2001).

Categorical regression may also be applied at a higher level, i.e. in an analysis of multiple studies (Hertzberg and Miller, 1985; Hertzberg, 1991; Hertzberg and Wymer, 1991). In this application of categorical regression severity categories are defined covering disparate endpoints. Most of these applications focus on estimating the likelihood that a given category of severity may occur at a given dose level.



**Figure 6.3** Dose-response model fitted to male (circles) and female (triangles) body weights plotted against log-dose. The CED (critical effect dose) associated with a CES (critical effect size) of 5% is estimated at 0.23 mg/kg, with lower confidence bound 0.21 mg/kg. The latter value can be considered as a BMDL. The plotted marks represent the (geometric) means of around 40 animals

### 6.4.3 Uncertainty

DRM should capture the relative uncertainties in the estimates of risk. This information will allow for the generation of confidence limits on health based guidance levels. However, such confidence levels will still capture only one part of the uncertainty inherent in these estimates. The risk manager will need to know what uncertainty is accounted for in the information

provided, and the risk assessment information will need to clearly indicate what uncertainty is not accounted for in a quantitative assessment.

One approach that has been used to capture variability in population response is calculation of population percentiles. Availability of dose response functions when linked with population based exposure assessments has allowed risk managers to calculate percentiles of populations above target exposure or intake level. Likewise, dose response functions have also been utilized to calculate percentiles of the population above target risk levels.

## **6.5 Explaining/interpreting the Output of the Dose-response Analysis**

Advice to the risk manager should describe the uncertainties inherent in such an approach to the use of DR data, such as uncertainties in the slope estimate in animals, the relevance of this slope to humans (such an approach is more appropriate if the response is a continuous variable, rather than quantal) and the appropriateness of the uncertainty factor applied to allow for species differences and human variability.

### **6.5.1 Outputs in the Observable Biological Range**

The output of the analysis takes the form of a numerical quantity - at present commonly a TDI or ADI derived from a NOAEL, which is a single point in the dose-response relationship. The DR analysis uses more of the available information by fitting a mathematical model to all the data in the observable biological range and then determining the dose associated with a specified response level. A statistical lower bound (for example, the 95% lower bound on the dose) is often used to account for statistical uncertainties (a BMDL) and for the level of health protection required by the risk manager. As with the NOAEL, the BMDL can be used as the starting point for deriving a health-based guidance value and/or margin of exposure (MOE). However, unlike the NOAEL, the BMD approach uses the whole range of experimental dose response data and therefore it is not limited by the doses selected by the investigators.

#### **6.5.1.1 Health Based Guidance Values**

On the basis of current practice, it appears that the BMD approach leads to doses that are usually quite similar to NOAELs for the studies in question (see Section 6.1.2). In the same way as for the derivation of the ADI/TDI, uncertainty factors, for example 100, are applied to the BMDL to obtain the health-based guidance value. However, the confidence intervals which are possible in the case of the BMD-derived health-based guidance value provide the risk manager with an increased understanding of the uncertainty associated with the risk assessment. This allows a more informed decision to be made when choosing between risk management options.

#### **6.5.1.2 Margin of Exposure (MOE)**

A MOE is determined by comparing the point of departure (the BMDL) to the actual or estimated human exposure. The MOE is used when limited toxicological or human data exist but the hazard identification and characterization data are insufficient to set a health-based guidance value. Alternatively, the MOE approach is used when it is inappropriate to derive a health-based guidance value due to the nature of the effect, such as for genotoxic carcinogenic compounds.

The acceptability of an MOE depends on its magnitude and is ultimately a risk management decision. To aid that decision, the risk assessor should provide information on the nature of the toxicity involved and nature and magnitude of the uncertainties, from both the toxicological and exposure perspectives. Although the risk assessor should not provide an assessment of the acceptability of the MOE, guidance on its adequacy taking into account the severity/nature of the toxicity, uncertainties and variability should be given, for example in terms of high, medium or low concern. The use of all the data by the DR analysis enables the uncertainties to be better defined. The MOE can also be used by the risk manager for priority setting.

There is no internationally accepted value for a MOE for a genotoxic and carcinogenic compound such that the exposure would not be a significant health risk. However, several institutions and countries have used the MOE approach and their conclusions provide examples of MOE values that have been considered acceptable:

- The National Health and Medical Research Council (NHMRC) in Australia concluded that a guideline dose for carcinogens present in soil could be calculated by application of uncertainty factors up to 50,000 to the BMD (not BMDL). The factor applied in any particular case would depend on the nature of the effects (NHMRC, 1999).
- The reciprocal of the MOE, the Exposure Potency Index (EPI), has been used by Health Canada for genotoxic and carcinogenic compounds in their Human Health Risk Assessment for Priority Substances under the Canadian Environmental Protection Act (Health Canada, 1994). MOE values of <5,000, 5,000-500,000 and >500,000 indicate high, medium and low priority, respectively.
- The Committee on Carcinogenicity (COC) in the UK recently considered derivation of the minimal risk level for a genotoxic and carcinogen compound. One proposal was that an adequate margin of exposure for carcinogenicity might be 10,000 (Gaylor, 1999; Gold et al., 2003). A particular carcinogenic impurity posed a negligible carcinogenic risk if an uncertainty factor of 10,000 was applied to the estimated 5% BMD (COC 2003). The MOE for average intakes for acrylamide in males in Norway has been estimated using the T25 value<sup>1</sup> and the LED10 (BMD 10%) methods. These approaches result in MOE values of 1306 and 1225 for T25 and LED10, respectively.

### 6.5.2 Outputs outside the Observable Biological Range

DRM evaluations can produce information in several formats including dose response functions that allow, along with estimates of exposure, for the prediction of risks at specified exposure levels and functions allowing for the estimation of exposure levels resulting in specified risks. In addition, DRM exercises can provide uncertainty analyses. The availability

---

<sup>1</sup> The tumorigenic descriptor T25 is the chronic daily dose which will give 25% of the animals tumours above background at a specific tissue site. The T25 is determined by linear interpolation from the lowest dose giving a statistically significant increase in tumours (Dybing et al., 1997)

of such outputs from DRM exercises can provide both opportunities for additional assessment, as well as challenges in interpretation for the risk manager.

Three different methods have been used or proposed for quantitative risk assessment by regulatory authorities in the US and Europe for non-threshold (genotoxic) carcinogens. In the area of food safety the US FDA has used a simple, direct method for low dose cancer risk assessment. A point on the dose-response curve is chosen below which the data no longer appear to be reliable (e.g., 1 to 10% tumour incidence) and a straight line is drawn from the upper confidence limit on risk at that point to the origin (Gaylor et al., 1997). The Linearised Multistage Model (LMS) was previously extensively used by the US EPA (1986). The LED10 method (BMDL) was later proposed by the US EPA (1996) and the T25 (Dybing et al., 1997, Sanner et al., 2001) method has been used in Europe (EC, 1999; SCCNFP, 2003). Lifetime cancer hazards may be calculated by linear extrapolation using LED10 and T25 as starting points, respectively. The results obtained with these extrapolation methods are in most cases nearly indistinguishable (Sanner et al., 2001). A measure for an assessment of concern may be arrived at by comparing the calculated risks for some specific human exposure scenario to such substances, with some default policy-determined risk level.

#### **6.5.2.1 Prediction of Risks at Specified Exposure Levels**

One type of output from DRM is the prediction of risks at specified exposure levels. This output can take the generic form of predicting “X number of health impacted individuals at exposure Y”. Examples of such estimates have been used to predict the number of excess lung cancer deaths due to smoking two packs of cigarettes per day, to predict the number of excess skin cancers from arsenic contaminated water and number of excess mortality cases due to air pollution. In the optimal case, such estimates are supported by parallel assessments that describe the uncertainty in such estimates by providing additional information on the range of estimates, rather than a single value. The risk manager can then make such statements such as “Up to X number of individuals may be impacted by exposure Y”. This same information can allow the risk manager to see how low the estimates of the health impact maybe, and when confidence limits are included in such estimates, many uncertain health impacts can be shown to include the potential for no health impacts. Assumptions inherent in such estimates that can impact risk manager interpretation include choice of models, choice of endpoints, limitations in initial datasets that were extrapolated.

One use of such information has been to evaluate the effect on risks of different maximum limits for a chemical. This type of consideration was included when JECFA evaluated aflatoxin B<sub>1</sub> and the impact of different maximum limits on risk. Similar assessments have also been performed for lead and fumonisins B<sub>1</sub> and B<sub>2</sub> (Carrington, et al., 1996; Humphreys et al., 2001). For example, the health impacts of current particulate standards (WHO, 2000b; WHO, 2003) have been estimated. Availability of such estimates can provide additional information for risk managers to conduct cost-benefit analyses, risk benefit assessments and evaluations of public health interventions.

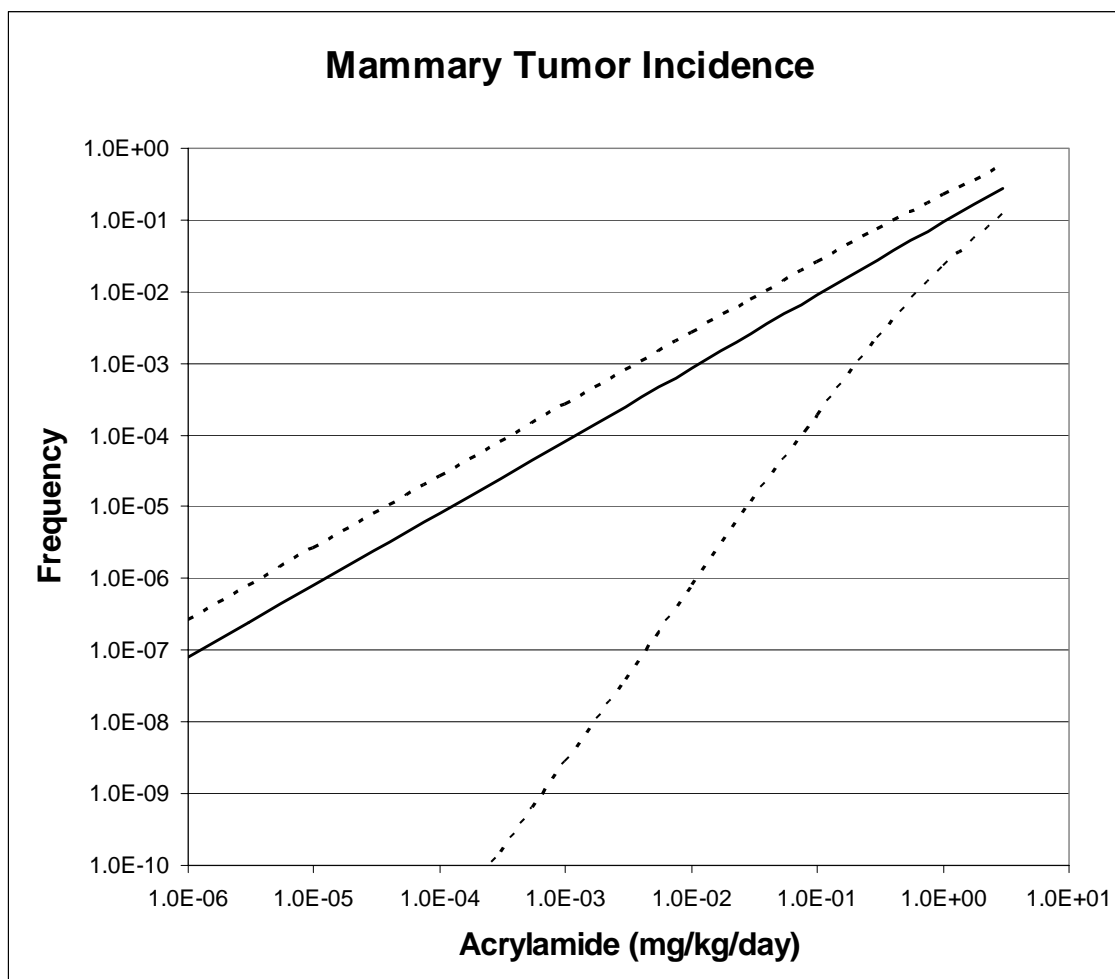
#### **6.5.2.2 Prediction of Exposure Levels Producing Specific Risk levels**

Another type of output from DRM is risk level estimates. In these estimates a specific level of risk is evaluated and the amount of exposure that would be estimated to result in that risk is determined. For example, a common level of risk related to carcinogen exposures that has been evaluated in the United States has been  $10^{-6}$  over a lifetime. Estimates of exposure that

would result in that level of risk have been determined and such estimates have been made for approximately 100 environmental pollutants (see "drinking water concentrations at specified risk levels" and "air concentrations at specified risk levels" at <http://www.epa.gov/iris>). For the risk manager, availability of such estimates can allow for development of risk based consistency in proposed regulatory actions.

### **6.5.2.3 Uncertainty Analyses**

A third type of output from DRM is that linked with uncertainty analysis. One example of such approaches is when the DRM output is linked with distributions population effects with confidence intervals. The result from such analysis is a distribution of potential population risks. For example, in Figure 6.4, the outputs for three models and two data sets were used to generate a set of 3000 different model parameters (2 data sets, 3 models, 500 bootstraps).



The correspondence between dose and effect at fixed frequency intervals are as follows

Frequency	1E-6	1E-5	1E-4	1E-3	1E-2	BMD10
Lower Bound	3.83E-06	3.83E-05	3.82E-04	3.82E-03	3.84E-02	4.03E-01
Median	1.20E-05	1.19E-04	1.18E-03	1.15E-02	1.10E-01	1.03E+00
Upper Bound	1.18E-02	3.03E-02	7.72E-02	2.11E-01	6.45E-01	2.64E+00

**Figure 6.4 Integrated uncertainty analysis for mammary tumors**

One approach that has been used to extrapolate dose response models beyond bioassay data has focused on the use of biomarker data to extend the dose response curve one to two orders of magnitude closer to environmentally relevant exposures. Such approaches can be facilitated when DRM data is available.

All these modelling approaches exhibit similar limitations and difficulties. A benefit is that DRM allows for the transfer of more quantitative toxicological data into risk manager assessment methods such as the cost benefit and risk benefits analyses. The limitation is the

question of whether the model outputs are accurate and representative of public health impacts.

## **6.6 Issues for Risk Managers**

### **6.6.1 Risk Assessment Issues**

#### **6.6.1.1 Population versus Individual Effects**

The potential health effect at the population level can be informed by the DRM. However, as the behaviour, environment, or biological characteristics may vary among individuals, a DRM may need to describe or model these characteristics to produce a prediction of adverse health effects in the population. The output of the DRM should identify the degree of any sub population effects.

#### **6.6.6.2 Risk Characterization**

The actual risk to the population of an adverse health effect requires consideration of both the likelihood and severity of the effect, as determined from the DRM when combined with the exposure to the chemical in the population under consideration. The exposure may be determined from consumption surveys, measurement of environmental media, direct contact information or biomarkers. (e.g., WHO, 2000a; Kroes et al., 2002)

Consideration of the DRM data together with exposure data will help identify populations at risk. This information, together with knowledge about the severity of the adverse health effects, will inform the risk management options.

### **6.7.2 Risk Management Issues**

#### **6.7.2.1 Risk Management Options**

A risk assessment may establish that a risk is of a sufficient magnitude that regulation or other type intervention may be warranted. DRM may then be used to evaluate the consequences of possible interventions that aim to reduce the risk. That is, a model may be used to estimate change in the likelihood of the adverse health effect occurring following implementation of a particular intervention. To date, alternative risk management options have been evaluated using DRM in a limited number of cases. For example, at the request of the Codex Committee on Food Additives and Contaminants, the 49<sup>th</sup> JECFA analysed the application of two hypothetical standards for aflatoxin contamination in food in model populations (WHO, 1999a).

However, a range of risk management interventions are available, with the types of interventions varying from a ban on a particular product (e.g., carcinogenic antibiotics, DDT), establishing regulatory limits (e.g., aflatoxin), advice on consumption or use patterns (e.g., consumption of predatory fish which accumulate high levels of methylmercury) and control at source of production (e.g., emissions of dioxins).

### 6.7.2.2 Cost and Risk Benefit Analyses

While health risk management decisions should be based on risk assessment, a number of other factors will inform the final decisions. In particular, it may also be necessary to undertake a cost benefit analysis (e.g., health costs to the community from exposure to aflatoxin versus the cost of implementation of a management strategy) and/or risk benefit (e.g. risk associated with methylmercury in fish versus nutritional benefits of fish consumption) and to assess the feasibility of the intervention, availability of alternatives and loss of products of economic value. These factors are beyond the scope of the assessment of the safety risks and will need to reflect wider societal factors.

### 6.7.2.3 Acceptable Level of Risk

Different institutions and countries may make different risk management decisions based on different perceptions of the risk that is deemed to be acceptable to society. The ADI, which usually incorporates a composite UF of 100 when based on animal studies, has been accepted by international institutions and countries as a health based guidance value. Although the DRM can give a prediction of the risk at various exposures, there is no international agreement on how to interpret this new information; the appropriate criteria for making or reversing risk decisions or the acceptable level of risk determined using this technique.

Some examples of risks considered acceptable by various countries and institutions are described in Sections 6.5.1 and 6.5.2. In all cases, the qualitative considerations will inform the final risk management decision.

A predicted risk, such as  $10^{-6}$ , determined from dose-response analysis has been used by some countries and institutions as posing no appreciable risk. Variations around the calculated risk by a factor of about 10 trigger further consideration of the qualitative aspects of the risk assessment, such as variability and uncertainty (Sanner et al., 2001; SCCNFP, 2003). In the case of compounds in drinking water considered to be genotoxic carcinogens, WHO has assigned guideline values associated with an estimated upper bound excess lifetime cancer risk of  $10^{-5}$  determined by a mathematical model (WHO, 1996). The US Occupational Safety and Health Administration (OSHA) has considered a lifetime cancer risk for workers higher than  $10^{-3}$  to represent an unacceptable high risk and its goal is to reduce this risk to less than  $10^{-5}$  (OSHA, 1983 and 1984). Proposals for the application of lifetime risk estimates in establishing tolerable risk levels have also been published in Europe (Bos et al., 2004).

## References

- Allen, BC, Kavlock RJ, Kimmel CA, Faustman EM (1994) Dose-response assessment for developmental toxicity: II. Comparison of generic benchmark dose estimates with NOAELs. *Fund. Appl. Toxicol.*, 23:487-495.
- Allen BC, Strong PL, Price CJ, Hubbard SA, Daston GP (1996), Benchmark Dose Analysis of Developmental Toxicity in Rats Exposed to Boric Acid. *Fundamental and Applied Toxicology* 32, 194-204.
- Akaike, H (1973) Information theory and an extension of the maximum likelihood principle. IN: *Proceedings of the Second International Symposium on Information Theory*, B. N. Petrov and F. Csaki, eds. Akademiai Kiado, Budapest, pp. 267-281.
- Barnes DG, Dourson ML (1988) Reference dose (RfD): Description and use in health risk assessments. *Regulatory Toxicology and Pharmacology* 8, 471-486.
- Barnes DG, Daston GP, Evans JS, Jarabek AM, Kavlock RJ, Kimmel CA, Park C, Spitzer HL (1995) Benchmark dose workshop: criteria for use of a benchmark dose to estimate a reference dose. *Regulatory Toxicology and Pharmacology* 21, 296-306.
- Bos PMJ, Baars BJ, van Raaij MTM. Risk assessment of peak exposure to genotoxic carcinogens: a pragmatic approach. *Toxicol Lett* 2004; 151: 43-50.
- Bozdogan H. (1987). Model-selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, 52, 345-370.
- Burnham KP and Anderson DR (2002) *Model Selection and Multimodel Inference*. Second Edition. Springer. New York
- Carrington, CD, Bolger PM, Scheuplein RJ (1996) Risk analysis of dietary lead exposure. *Food Additives Contaminants* 13:61-76.
- Carrington, CD and Bolger PM (2000). A Pooled Analysis of the Iraqi and Seychelles Methylmercury Studies. *Human Ecological Risk Assessment* 6:323-340.
- Codex Alimentarius Commission (2003) Report of the 26<sup>th</sup> Session of the Codex Alimentarius Commission, ALINORM 03/41, Appendix IV [http://www.fao.org/codex/alinorm03/a103\\_41e.pdf](http://www.fao.org/codex/alinorm03/a103_41e.pdf).
- Cogliano VJ, Baan RA, Straif K, Grosse Y, Secretan MB, El Ghissassi F, Kleihues P. (2004) The science and practice of carcinogen identification and evaluation. *Environ Health Perspect.*, 112(13):1269-74. <http://ehp.niehs.nih.gov/docs/2004/6950/abstract.html>
- Crump KS (1984) A new method for determining allowable daily intakes. *Fundamental and Applied Toxicology* 4, 854-871.
- Crump KS (1995) Calculation of Benchmark Doses from Continuous Data, *Risk Analysis* 15(1): 79-89.

- Crump K (2002). Critical issues in benchmark calculations from continuous data, Crit Rev Toxicol 32(3): 133-53.
- Davidian M, Giltinan DM (1995) Nonlinear Models for Repeated Measurement Data. Chapman and Hall, London.
- Dourson ML, Stara JF (1983) Regulatory history and experimental support of uncertainty (safety) factors. Regulatory Toxicology and Pharmacology, 2, 224-238.
- Dourson ML, Hertzberg RC, Hartung R, Blackburn K (1985) Novel approaches for the estimation of acceptable daily intake, Toxicology and Industrial Health 1, 23-41.
- Committee on Carcinogenicity (2003) Minutes of UK Committee on Carcinogenicity of Chemicals in Food, Consumer Products and the Environment meeting 26 June 2003 <http://www.advisorybodies.doh.gov.uk/coc/meetings/coc032.htm>
- Dybing E, Sanner T, Roelfzema H, Kroese D, Tennant RW. (1997) A simplified carcinogenic potency index. Description of the system and study of correlations between carcinogenic potency and species/site specificity and mutagenicity. Pharmacol Toxicol; 80: 272-279.
- Dybing E, Sanner T (2003) Risk assessment for acrylamide in food. Toxicological Sciences 75:7-15.
- European Commission (1999) European Commission Guidelines for setting specific concentration limits for carcinogens in Annex 1 of Directive 67/548/EEC. Inclusion of potency considerations. Commission Working Group on the Classification and Labelling of Dangerous Substances. Brussels, 1999.
- European Commission (2002) White Paper on Food Safety. 12 January 2002, Comm (1999)719final.
- Edler LK, Dourson M, Kleiner J, Mileson B, Nordmann H, Renwick A, Slob W, Walton K., Würtzen G (2002) Mathematical modelling and quantitative methods, Food and Chemical Toxicology 40, 283-326.
- Edler LK, Dourson M, Kleiner J, Mileson B, Nordmann H, Renwick A, Slob W, Walton K., Würtzen G (2002) Mathematical modelling and quantitative methods, Food and Chemical Toxicology 40, 283-326
- Efron B (1987) Better bootstrap confidence intervals. Journal of the American Statistical Society 82: 171 – 185.
- Efron B, Tibshirani, R (1993) An Introduction to the Bootstrap. Chapman and Hall. London.
- Evans JS, Graham JD, Gray GM, Sielken RL (1994), ‘A distributional approach to characterizing low-dose cancer risk’, Risk Anal 14, 25-34.
- Evans M, Hastings N, and Peacock B (1993). Statistical Distributions, 2<sup>nd</sup> Ed. John Wiley & Sons, New York.

- Fitzgerald DJ, Robinson NI and Pester, BA (2004). Application of Benzo(a)pyrene and Coal Tar Tumor Dose-Response Data to a Modified Benchmark Dose Method of Guideline Development. *Environ Health Perspect* 112:1341-46/
- Fowles, JR, Alexeeff GV, Dodge D (1999) The use of benchmark dose methodology with acute inhalation lethality data. *Regulatory Toxicology and Pharmacology* 29: 262-278.
- Garvin, T. (2001) Analytical Paradigms: The Epistemological Distances between Scientists, Policy Makers, and the Public. *Risk Anal*, 21(3), 443-455.
- Gaylor, DW, Axelrad, JA, Brown, RP, Cavagnaro, JA, Cyr, WH, Hulebak, KI, Lorentzen, RJ, Miller, MA, Mulligan, LT and Schwetz, BA (1997) Health Risk Assessment Practices in the U.S. Food and Drug Administration. *Reg. Tox. Pharm.* 26:307-321.
- Gaylor DW, Kodell RL Chen JJ, Krewski D (1999) A unified approach to risk assessment for cancer and noncancer endpoints based on benchmark doses and uncertainty/safety factors. *Regulatory Toxicology and Pharmacology* 29: 151-157.
- Gaylor D, Aylward L (2004). "An evaluation of benchmark dose methodology for non-cancer continuous-data health effects in animals due to exposures to dioxin (TCDD)." *Regulatory Toxicology and Pharmacology* 40: 9-17.
- Gaylor, D and Slikker W (1994). "Modelling for risk assessment of neurotoxic effects." *Risk Anal* 14(3): 333-8.
- Gelman A, Carlin JB, Stern HS, and Rubin DB (2004). *Bayesian Data Analysis: Second Edition*. Chapman and Hall/CRC. Boca Raton.
- Ghani AC, Ferguson, N.M, Donnelly,CA. & Anderson, R. M. Predicted vCJD mortality in Great Britain. *Nature* 406, 583-584 (2000)
- Gold LS, Gaylor DW, Slone TH (2003) Comparison of cancer risk estimates based on a variety of risk assessment methodologies, *Regulatory Toxicology and Pharmacology* 237:45-53.
- Hacking I (1976) The Great Decision. In: The Emergence of Probability, Cambridge University Press, Cambridge, pp 63-72.
- Hasselblad V, Jarabek AM (1995) Dose-Response Analysis of Toxic Chemicals. Bayesian Biostatistics. D. A. Berry and D. K. Stangl. New York, Marcel Dekker, Inc.
- Hertzberg RC (1991) Quantitative extrapolation of toxicological finds. In: Krewski, D, Franklin C (Eds.) *Statistics in Toxicology*, Gordon & Breach Science, New York, NY.
- Hertzberg RC Miller M (1985) A statistical model for species extrapolation using categorical response data. *Toxicology and Industrial Health*, 43-57.
- Hertzberg RC and Wymer L (1991) Modelling the severity of toxic effects. *Air & Waste Management Assoc. Proceedings Papers 84<sup>th</sup> Ann. Meeting & Exhibition*, 16-21 June 1991, British Columbia.

- Humphreys SH, Carrington CD, Bolger PM (2001) A quantitative risk assessment for Fumonisin B1 and B2 in corn. *Food Additives Comtam.* 18:211-220.
- Kim AH, Kohn, MC, Portier CJ, Walker NJ (2002) Impact of physiologically based pharmacokinetic modeling on benchmark dose calculations for TCDD-induced biochemical responses. *Regul Toxicol Pharmacol* 36, 287-96
- Kimmel CA, Gaylor DW (1988) Issues in qualitative and quantitative risk analysis for development toxicology. *Risk Anal* 8, 15-20.
- Kodell RL, Chen JJ et al. (1995). "Neurotoxicity modelling for risk assessment." Regul Toxicol Pharmacol 22(1): 24-9.
- Kroes R, Müller D, Lambe J, Löwik MRH, van Klavern J, Kleiner J, Massey R, Mayer S, Urieta I, Verger P, Visconti A (2002) Assessment of intake from the diet, *Food and Chemical Tox.*, 40:327-385.
- Larson JL; Wolf DC, Mery, S, et al. (1995) Toxicity and cell proliferation in the liver, kidney and nasal passages of female F344 rats induced by chloroform administered by gavage. *Food Chem Toxicol* 33:443-45
- Lehman AJ, Fitzhugh OG (1954) Ten-fold safety factor studies. *Assoc Food Drug Officials of the US Quarterly Bulletin XVIII* (1): 33-35
- Moerbeek M, Piersma AH, et al. (2004). "A comparison of three methods for calculating confidence intervals for the benchmark dose." Risk Anal 24(1): 31-40.
- Murrell, JA, Portier, CJ, and Morris, RW (1998) Characterizing dose-response: I: Critical assessment of the benchmark dose concept, *Risk Analysis* 18: 13-26.
- National Health and Medical Research Council (1999) Toxicity Assessment for Carcinogenic Soil Contaminants. Commonwealth of Australia. Canberra.
- National Research Council (1983) Risk Assessment in the Federal Government: Managing the Process, NAS Press, Washington, DC, USA.
- National Research Council (1994) Science and Judgement In Risk Assessment, NAS Press, Washington, DC, USA
- National Research Council (1996) Understanding Risk: Informing Decisions in a Democratic Society. NAS Press, Washington, DC, USA
- NTP (2002) Report on Carcinogens. 10<sup>th</sup> ed. Research Triangle Park, NC: National Toxicology Program.  
<http://http-server.niehs.nih.gov/NewHomeRoc/AboutRoc.html>
- OSHA (1983) Occupational Exposure to Inorganic Arsenic: Supplemental Statement of Reasons for the Final Rule, 48 F.R. 1,864.

OSHA (1984) Occupational Exposure to Ethylene Oxide: Final Standard, 49 F.R. 46,936 (1984).

Piersma AH, Verhoef A, te Biesebeek JD, Pieters MN, Slob W (2000) Developmental toxicity of butyl benzyl phthalate in the rat using a multiple dose study design *Reproductive Toxicology* 14: 417-425

Pieters MN, Bakker M, and Slob W. (2004). Reduced intake of deoxynivalenol in The Netherlands: a risk assessment update. *Toxicology letters*, 153: 145-153.

Portier CJ (1994) Biostatistical issues in the design and analysis of animal carcinogenicity experiments. *Environ Health Perspect*, 102 Suppl 1, 5-8.

Portier C, Kohn, M (1996). A biologically-based model for the carcinogenic effects of 2378-TCDD in female Sprague-Dawley rats. *Organohalogen Compounds* 29, 222-227.

Presidential Commission (1997) Risk Assessment and Risk Management in Regulatory Decision-Making, Vol. 2, The Presidential/Congressional Commission on Risk Assessment and Risk Management. Washington, DC, USA.

R Development Core Team (2004). R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing.

Raftery A, Madigan D, Hoeting J (1997). Bayesian model averaging for linear regression models. *Journal of the American Statistical Association*, 92:179-191.

Rescher N (1969). *Many-Valued Logic*. McGraw-Hill, New York:

Renwick AG, Barlow, SM, Hertz-Picciotto I, Boobis AR, Dybing E, Edler L, Eisenbrand G, Grieg, JB, Kleiner J, Lambe J, Müller DJG, Smith MR, Tritscher A, Tuijtelaars S, van den Brandt PA, Walker R, and Kroes R (2003) Risk characterization of chemicals in food and diet, *Food Chem. Tox.*, 41:1211-1271.

Renwick AG, Lazarus NR (1998) Human variability and noncancer risk assessment-an analysis of the default uncertainty factor. *Regulatory Toxicology and Pharmacology*, 27, 3-20.

Sanner T, Dybing E, Willems MI, Kroese ED. (2001) A simple method for quantitative risk assessment of non-threshold carcinogens based on the dose descriptor T25. *Pharmacol. Toxicol.* 2001; 88: 331-341.

SCCNFP (2003). The Scientific Committee on Cosmetic Products and Non-Food Products Intended for Consumers. The SCCNFP's notes of guidance for the testing of cosmetic ingredients and their safety evaluation. 5<sup>th</sup> Revision. Adopted by the SCCNFP during the 25<sup>th</sup> plenary meeting of 20 October 2003, Brussels.

Slob W (2004). PROAST: A software package for dose-response modelling, deriving a benchmark dose and probabilistic hazard characterization. Bithoven, RIVM.

Slob W, Pieters MN (1998) A probabilistic approach for deriving acceptable human intake limits and human health risks from toxicological studies: general framework. *Risk Analysis* 18, 787-798.

Slob W, Moerbeek M, Rauniomaa E, Piersma AH (2005) A statistical evaluation of toxicity designs for the estimation of the benchmark dose in continuous endpoints. *Toxicological Sciences*, 84, 167–185.

Thompson KM (2002). Variability and uncertainty meet risk management and risk communication *Risk Analysis*. 22(3):647-654.

Timchalk C, Nolan RJ, Mendrala AL, Dittenber DA, Brzak, KA, Mattsson, JL (2002). A physiologically based pharmacokinetic and pharmacodynamic (PBPK/PD) model for the organophosphate pesticide chlorpyrifos in rats and humans. *Toxicological Sciences* 66, 34 – 53.

USEPA (1986) US Environmental Protection Agency. Guidelines for carcinogen risk assessment. *Fed Reg* 1986; 51: 33992-34003.

USEPA (1996) US Environmental Protection Agency. Proposed guidelines for carcinogen risk assessment. *Fed Ref* 1996; 61: 17960-180011.

USEPA (2001). Help Manual for Benchmark Dose Software. EPA 600/R-00/014F

USEPA (2003) Draft Final Guidelines for Carcinogen Risk Assessment. EPA/630/P-03/0001A. Washington DC.

<http://www.epa.gov/ncea/raf/cancer/2003.htm>

Williams GM, Iatropoulos MJ (2002) Alteration of liver cell function and proliferation: differentiation between adaptation and toxicity. *Toxicol Pathol* 30(1), 41-53.

WHO (1987) Principles for the Safety Assessment of Food Additives and Contaminants in Food, *Environmental Health Criteria* 70, International Programme on Chemical Safety, Geneva.

WHO (1994) Assessing Human Health Risks of Chemicals: Derivation of Guidance Values for Health-based Exposure Limits. *Environmental Health Criteria* 170, International Programme on Chemical Safety, Geneva.

WHO (1996) Guidelines for drinking-water quality. Second Edition, Vol. 2. Health criteria and other supporting information, Geneva.

WHO (1997) Food consumption and exposure assessment of chemicals, Report of a FAO/WHO Consultation, 10-14 February 1997, WHO/FSF/97.5. Available from Department of Food Safety, World Health Organization, 20 Avenue Appia, CH-1211 Geneva 27, Switzerland.

WHO (1999) Principles for the Assessment of Risks to Human Health for Exposure to Chemicals, *Environmental Health Criteria* 210, International Programme on Chemical Safety, Geneva.

WHO (1999a) Evaluation of Certain Food Additives and Contaminants, Forty-ninth report of the Joint FAO/WHO Expert Committee on Food Additives, Geneva.

WHO (2000a) Human Exposure Assessment, Environmental Health Criteria 214, International Programme on Chemical Safety, Geneva.

WHO (2000b) Air Quality Guidelines for Europe, 2nd edition. WHO Regional Office for Europe, Copenhagen.

WHO (2000c) Lead. In: Safety evaluation of certain food additives and contaminants, WHO Food Additive Series 44. World Health Organization, Geneva, pp . 273-312.

WHO (2001) Guidance document for the use of chemical-specific adjustment factors (CSAFs) for interspecies differences and human variability in dose-concentration response assessment. International Programme on Chemical Safety, Geneva. <http://www.ipcsharmonize.org/CSAFsummary.htm>.

WHO (2002) Evaluation of Certain Food Additives and Contaminants, Fifty-seventh report of the Joint FAO/WHO Expert Committee on Food Additives, Technical Report Series 909, Geneva.

WHO (2003) Health Aspects of Air Pollution with Particulate Matter, Ozone and Nitrogen Dioxide. Report of a Working Group. WHO Regional Office for Europe,

WHO (2004a) Risk assessment of *Listeria monocytogenes* in ready-to-eat foods, Microbiological Risk Assessment Series, No. 5, Technical Report, Geneva

WHO (2004b) IPCS Risk Assessment Terminology, International Programme on Chemical Safety, Geneva.

Woutersen, R.A., Jonker, D., Stevenson, H., J.D. te Biesebeek, and Slob, W. (2001). The benchmark approach applied to a 28-day toxicity study with Rhodorsil Silane in rats: the impact of increasing the number of dose groups. *Food and Chemical Toxicology*, 39: 697-707.

## Appendix I: Terminology

**Acceptable Daily Intake/Tolerable Daily Intake/Reference Dose:** estimated maximum amount of an agent, expressed on a body mass basis, to which an individual in a (sub) population may be exposed daily over its lifetime without appreciable health risk.

**Acceptable Risk:** a risk management term. The acceptability of risk depends on scientific data, social, economic, and political factors, and on the perceived benefits arising from exposure to an agent.

**Additional Risk (Extra Risk):** the additional proportion of total animals that respond in the presence of the dose, or the probability of response at dose  $d$ ,  $P(d)$ , minus the probability of response in the absence of exposure,  $P(0)$ .

**Adverse Effect:** change in the morphology, physiology, growth, development, reproduction or life span of an organism, system, or (sub) population that results in an impairment of functional capacity, an impairment of the capacity to compensate for additional stress, or an increase in susceptibility to other influences.

**Akaike Information Criteria:** a statistical procedure that provides a measure of the goodness-of-fit of a dose-response model to a set of data.

**Assessment Factor:** numerical adjustment to extrapolate from experimentally determined (dose-response) relationships to estimate the agent exposure below which an adverse effect is not likely to occur.

**Asymptotic Test:** statistical tests that approach known properties as sample sizes increase.

**Benchmark Concentration (BMC):** the concentration of a substance that is associated with a specified low incidence of risk of a health effect, or the concentration associated with a specified measure or change of a biological effect.

**Benchmark Dose (BMD):** an exposure due to a dose of a substance associated with a specified low incidence of risk, generally in the range of 1% to 10%, of a health effect; or the dose associated with a specified measure or change of a biological effect.

**Benchmark Dose Lower Confidence Limit (BMDL):** a lower one-sided confidence limit on the BMD.

**Benchmark Response (BMR):** the response, generally expressed as in excess of background, at which a benchmark dose or concentration is desired.

**Bernoulli Distribution:** a theoretical distribution of the number of successes in a finite set of independent trials with a constant probability of success. It is discrete distribution having two possible outcomes labelled by  $n = 0$  and  $n = 1$  in which  $n = 1$  ("success") occurs with probability  $p$  and  $n = 0$  ("failure") occurs with probability  $q \equiv 1 - p$ , where  $0 < p < 1$ .

**Beta-Binomial Distribution:** a statistical distribution of clustered values, e.g., measures on offspring in a litter, where the average proportions of an event for clusters are described by a

Beta distribution and the proportions of events in a cluster are described by a binomial distribution.

Binomial Distribution: the statistical distribution of the probabilities of observing 0,1,2,- - -,n events in a sample of n independent trials each with the same individual probability that the event occurs.

Bootstrap: a statistical technique based on multiple re-sampling with replacement of the sample values or re-sampling of estimated distributions of the sample values that is used to calculate confidence limits or perform statistical tests for complex situations or where the distribution of an estimate or test statistic cannot be assumed.

Cancer Potency (Cancer Slope Factor): a number that estimates the cancer risk (incidence) for a lifetime exposure to a substance per unit of dose which is generally expressed as mg / kg body wt / day.

Categorical Data: results obtained where observations or measurements on individuals or samples are stratified according to degree or severity of an effect, e.g., none, mild, moderate, or severe.

Categorical default factor: a factor based on common characteristics of a group of compounds, e.g., physical/chemical properties or pathways of metabolism.

Chemical Specific Adjustment Factor: a factor based on quantitative chemical-specific toxicokinetic or toxicodynamic data, which replaces some or all of the default uncertainty factor.

Chi-square Test: a statistical test used to examine the deviation of an observed number of events from an expected number of events.

Clustered Data: measurements collected on some grouping of individuals, e.g., litters in reproductive and developmental studies.

Confidence Interval (Two-Sided): an estimated interval from the lower to upper confidence limit of an estimate of a parameter. This interval is expected to include the true value of the parameter with a specified confidence percentage, e.g., 95% of such intervals are expected to include the true values of the estimated parameters.

Confidence Interval (One-Sided ): an interval below the estimated upper confidence limit, or interval above the estimated lower confidence limit, that is expected to include the true value of an estimated parameter with a specified confidence ( percent of the time).

Confidence Limit: an estimated value below (or above) which the true value of an estimated parameter is expected to lie for a specified percentage of such estimated limits.

Constrained Dose-Response Model: estimates of one or more parameters of the model restricted to a specified range, e.g., equal to or greater than zero.

**Continuous Data:** effects Measured on a continuum, e.g., organ weight or enzyme concentration, as opposed to quantal or categorical data where effects are classified by assignment to a class.

**Convergence:** estimates of a parameter approach a single value with increasing sample size or increasing number of computer iterations.

**Convex:** region of a dose-response relationship that curves upward, i.e., the slope becomes steeper with increasing dose.

**Correlated Binomial Distribution:** clustered data where the individual values in a cluster, e.g., a litter, each have the same probability of expressing an effect.

**Covariate:** an independent variable other than dose that may influence the outcome of an effect, e.g., age, body weight, or polymorphism.

**Critical effect:** the adverse effect, or its known precursor, that is relevant to human risk assessment and that occurs in the dose/concentration scale in the most sensitive animal species.

**Cubic:** an effect is a function of a measure raised to the third power.

**Degrees of Freedom:** for dose-response model fitting, the number of data points minus the number of model parameters estimated from the data.

**Default value:** pragmatic, fixed or standard value used in the absence of relevant data.

**Delta Method:** variance of a function of random variables approximated from the derivatives of the function with respect to the random variables and the variances of the random variables.

**Dichotomous Data:** quantal data where an effect for an individual may be classified by one of two possibilities, e.g., dead or alive, with or without a specific type of tumour.

**Dispersion:** variation (differences) from a central (mean or median) value.

**Dose:** total amount of an agent administered to, taken up or absorbed by an organism, system or (sub) population.

**Dose-Response:** relationship between the amount of an agent administered to, taken up or absorbed by an organism, system or (sub) population and the change developed in that organism, system or (sub) population in reaction to the agent.

**Dose-Response Assessment:** analysis of the relationship between the total amount of an agent administered to, taken up or absorbed by an organism, system or (sub)population and the changes developed in that organism, system or (sub)population in reaction to that agent, and inferences derived from such an analysis with respect to the entire population.

**Dose-Response Model:** a mathematical relationship (function) that relates (predicts) a measure of an effect to a dose.

**Dose-Response Trend:** relationship between incidence or severity of a biological effect and a function of dose. Simply the slope for a linear dose-response.

**ED<sub>x</sub>:** effective dose associated with a biological effect in x% of the individuals. Dose may be the external exposure often expressed in mg per day of the substance per kg body weight raised to a power (generally 1, 3/4, or 2/3) or area under the curve (AUC) in blood or target tissue 20 where the substance remains in the body over a period of time.

**Estimate:** an empirical value derived from data for a parameter.

**Exposure:** concentration or amount of a particular agent that reaches a target organism, system or (sub) population in a specific frequency for a defined duration.

**Excess Risk:** proportion of individuals or animals observed or estimated to possess an effect in addition to the spontaneous background risk.

**Gamma Distribution:** a unimodal statistical distribution ( relative proportion of responders as a function of some measure ) that is restricted to effects greater than or equal to zero that can describe a wide variety of shapes, e.g., flat, peaked, asymmetrical.

**Gaussian (Normal) Distribution:** a unimodal symmetrical (bell-shaped) distribution where the most prevalent value is the mean (average) and the spread is measured by the standard deviation. Mathematically, the distribution varies from minus infinity with zero probability to plus infinity with zero probability.

**Generalized Estimating Equation (GEE):** a statistical technique used for estimating parameters that requires only specification of the first two moments of the distribution of the estimator as opposed to a complete specification of the distribution.

**Goodness-of-Fit:** a statistic that measures the dispersion of data about a dose-response curve in order to provide a test for rejection of a model due to lack of an adequate fit, e.g., a P-value < 0.1.

**Hazard Identification:** the identification of the type and nature of adverse effects that an agent has as inherent capacity to cause in an organism, system or (sub) population.

**Hill Equation:** a dose-response curve, frequently used for enzyme kinetics, that monotonically approaches an asymptote (maximum value) as a function of dose raised to a power.

**Hybrid Model:** for continuous data establishes abnormal values based on the extremes in controls (unexposed individuals or animals) and estimates the risk of abnormal levels as a function of dose.

**Incidence:** proportion or probability of individuals or animals exhibiting an effect that varies from zero to one, sometimes expressed as a percent from 0% to 100%.

**Independence:** the result in one animal or individual does not influence the result in another animal or individual.

**Intercept Term:** the estimated value at zero dose or the dose corresponding to a zero effect.

**Least Squares:** a statistical procedure that estimates the values of dose-response parameters such that the sum of squares of deviations of data points from their estimated values is minimized, i.e., minimizes the estimated variance.

**Likelihood Ratio:** ratio of the probability that the observed data arise from a set of model parameters relative to the maximum probability that arises from the set of maximum likelihood estimates.

**Linear Dose-Response Model:** the amount of change in a response is proportional to the amount of change in some function of dose.

**Linearized Multistage Model:** dose-response model based on the multistage model of carcinogenesis that is restricted to a form that is approximately linear at low doses.

**Local Maximum:** mathematical solution that maximizes a function in a region that may not be the overall global maximum.

**Likelihood Function:** relative probabilities that various values of population parameters would arise from the sample observations.

**Logistic Model:** a sigmoid (S-shaped) function that relates the proportion of individuals with a specified characteristic to an independent variable.. random variable has a normal distribution

**Lognormal Distribution:** a mathematical description where the natural logarithm of a random variable has a normal distribution

**Log Transformation:** logarithm of raw data.

**Lowest-observed-adverse-effect level (LOAEL):** the lowest concentration or amount of a substance, found by experiment or observation, that causes an adverse alteration of morphology, functional capacity, growth, development or life span of the target organisms distinguishable from normal (control) organisms of the same species and strain under the same defined conditions of exposure.

**Lowest-observed-effect level (LOEL):** the lowest concentration or amount of a substance, found by experiment or observation, that causes any alteration of morphology, functional capacity, growth, development or life span of the target organisms distinguishable from normal (control) organisms of the same species and strain under the same defined conditions of exposure.

**Maximum Likelihood Estimate (MLE):** estimate of a population parameter most likely to have produced the sample observations.

**Mechanism of action:** a detailed description of the precise chain of events from the molecular level to gross macroscopic or histopathological toxicity.

**Michaelis-Menten Equation:** a dose-response curve, frequently used for enzyme kinetics, with maximum slope at zero dose that approaches a maximum asymptote at increasing dose.

Margin of Exposure (MOE): ratio of the no-observed-adverse-effect level (NOAEL) for the critical effect to the theoretical, predicted or estimated exposure dose or concentration

Margin of Safety (MOS): margin between the reference dose and the actual exposure dose or concentration

Mode of action: a series of events that may lead to induction of the relevant end-point of toxicity for which the weight of evidence supports plausibility.

Monotonic Dose-Response: a dose-response that never decreases as dose increases. A monotonic function may be flat (constant) up to a threshold dose or may be flat at high doses if a biological limit, e.g., saturation, is attained.

Multinomial: animals or individuals may be classified by more than two (binomial) categories, e.g., in a reproductive study fetuses may be: dead, alive normal, or alive abnormal.

Negligible Risk: a risk management term. In cases where a quantitative risk estimate has been made, it is any risk less than an upper-bound incremental lifetime risk calculated using conservative risk assessment techniques such as the BMD.

No-Observed-Adverse-Effect Level (NOAEL): the highest concentration or amount of a substance, found by experiment or observation, that causes no detectable adverse alteration of morphology, functional capacity, growth, development or life span of the target organisms under defined conditions of exposure.

Nonlinear Dose-Response Model: mathematical relationship that cannot be expressed simply as the change in response being proportional to the amount of change of some function of dose.

Normal Distribution: a mathematical description where a continuous random variable  $x$  with a mean  $\mu$  and a variance  $\sigma^2$  has a probability density function

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}$$

Objective Function: choice of function that is optimized for maximum likelihood estimation.

Ordinal Data: integers designating the rank, order, or counts.

P-Value: in testing a hypothesis, the probability of a type I error (false positive). The probability that the sample (experimental) results are compatible with a specific hypothesis.

Parameter: a value used to numerically describe a population of values, e.g., the mean and standard deviation; or a value used to describe a dose-response curve, e.g., the intercept and the slope of a linear dose-response.

Point of Departure (POD): the point on a dose-response curve established from experimental data, e.g., the benchmark dose, generally corresponding to an estimated low effect level (e.g., 1% to 10% incidence of an effect). Depending on the mode of action and available data, some

form of extrapolation below the POD may be employed for low-dose risk assessment or the POD may be divided by a series of uncertainty factors to arrive at a reference dose.

Polynomial: a mathematical function of the sum of a constant, linear term, quadratic term, cubic term, etc.

Probability: the proportion (on a scale of 0 to 1) of cases for which a particular event occurs. Zero indicates the event never occurs and one indicates the event always occurs.

Probability Distribution: a mathematical description of the relative probabilities of all possible outcomes of a measurement.

Probit Function: assumes that the relative probabilities of effects as a function of dose are described by a Normal distribution. The cumulative probability as a function of dose has a sigmoid shape.

Profile Likelihood: a plot of the likelihood function versus the estimated value of a parameter.

Quadratic Term: a quantity in a mathematical formula that is raised to the second power (squared).

Quantal Data: dichotomous (Binomial) classification where an individual or animal is placed in one of two categories, e.g., dead or alive, with or without a particular type of tumour, normal or abnormal level of a hormone.

Quantile: percentile (cumulative probability) of a distribution that ranges from zero to the 100th percentile.

Quasi-Likelihood: likelihood function that is not totally defined and generally based on only an expression including the mean and variance.

Rectangular Hyperbola: a mathematical function of the form  $y^2 = x^2 + c$ , where  $x$  and  $y$  are variables and  $c$  is a constant.

Regression Analysis: a statistical process that produces a mathematical function (regression equation) that relates a dependent variable (biological effect) to independent variable, e.g., dose rate, duration of exposure, age.

Repeated Measures: a biological endpoint is measured for the same individual or animal at different times (ages).

Response: change developed in the state or dynamics of an organism, system or (sub) population in reaction to exposure to an agent.

Residual Variance: the variance in experimental measurements remaining after accounting for the variance due to the independent variables, e.g., dose rate, duration of exposure, age. Typically referred to as the inherent unaccountable experimental variation.

Residuals: the numerical differences between observed and estimated effects.

**Risk:** the probability of an adverse effect in an organism, system or (sub)population caused under specified circumstances by exposure to an agent.

**Risk Assessment:** a process intended to calculate or estimate the risk to a given target organism, system or (sub)population, including the identification of attendant uncertainties, following exposure to a particular agent, taking into account the inherent characteristics of the agent of concern as well as the characteristics of the specific target system.

**Risk Characterization:** the qualitative and, wherever possible, quantitative determination, including attendant uncertainties, of the probability of occurrence of known and potential adverse effects of an agent in a given organism, system or (sub)population, under defined exposure conditions.

**Safety Factor:** composite (reductive) factor by which an observed or estimated no-observed-adverse effect level (NOAEL) is divided to arrive at a criterion or standard that is considered safe or without appreciable risk.. (See uncertainty factor.)

**Severity:** the degree to which an effect changes and impairs the functional capacity of an organ system.

**Shape Parameter:** the exponent on dose in a dose-response function that dictates the curvature of the function.

**Threshold:** dose or exposure concentration of an agent below that a stated effect is not observed or expected to occur.

**Threshold of Toxicological Concern:** an exposure threshold value below which there is a very low probability of an appreciable risk to human health

**Toxicodynamics:** the process of interaction of chemical substances with target sites and the subsequent reactions leading to adverse effects.

**Toxicokinetics:** the process of the uptake of potentially toxic substances by the body, the biotransformation they undergo, the distribution of the substances and their metabolites in the tissues, and the elimination of the substances and their metabolites from the body. Both the amounts and the concentrations of the substances and their metabolites are studied. The term has essentially the same meaning as pharmacokinetics, but the latter term should be restricted to the study of pharmaceutical substances.

**Uncertainty:** imperfect knowledge concerning the present or future state of an organism, system or (sub)population under consideration

**Uncertainty Factor:** reductive factor by which an observed or estimated no-observed-adverse effect level (NOAEL) is divided to arrive at a criterion or standard that is considered safe or without appreciable risk. see Safety Factor

**Unconstrained Dose-Response Model:** no restrictions imposed on the estimates of parameters.

**Upper-Tail Probability:** probability that a variable exceeds a specified value.

Validation: process by which the reliability and relevance of a particular approach, method, process or assessment is established for a defined purpose.

Variability: observable diversity in biological sensitivity or response, and in exposure parameters.

Variance: measure of variability, standard deviation squared.

Weibull: form of a dose-response curve characterized by a relatively shallow slope at low doses that increases sharply as dose increases before leveling off at high doses.

Weighted Least Squares Estimate: parameter estimate obtained by minimizing the sum of squares of observed and estimated values weighted by a function, frequently the reciprocal of the variance of an observation.