参考資料９

IPCS

INTERNATIONAL PROGRAMME ON CHEMICAL SAFETY

WHO ILO UNEP

Environmental Health Criteria 239

Principles for Modelling Dose–Response for the Risk Assessment of Chemicals

IOMC INTER-ORGANIZATION PROGRAMME FOR THE SOUND MANAGEMENT OF CHEMICALS
A cooperative agreement among UNEP, ILO, FAO, WHO, UNIDO, UNITAR and OECD

World Health Organization

This report contains the collective views of an international group of experts and does not necessarily represent the decisions or the stated policy of the United Nations Environment Programme, the International Labour Organization or the World Health Organization.

Environmental Health Criteria 239

PRINCIPLES FOR MODELLING DOSE–RESPONSE FOR THE RISK ASSESSMENT OF CHEMICALS

First draft prepared by the WHO Task Group on Environmental Health Criteria on Principles for Modelling Dose–Response for the Risk Assessment of Chemicals

Published under the joint sponsorship of the United Nations Environment Programme, the International Labour Organization and the World Health Organization, and produced within the framework of the Inter-Organization Programme for the Sound Management of Chemicals.

The **International Programme on Chemical Safety (IPCS)**, established in 1980, is a joint venture of the United Nations Environment Programme (UNEP), the International Labour Organization (ILO) and the World Health Organization (WHO). The overall objectives of the IPCS are to establish the scientific basis for assessment of the risk to human health and the environment from exposure to chemicals, through international peer review processes, as a prerequisite for the promotion of chemical safety, and to provide technical assistance in strengthening national capacities for the sound management of chemicals.

The **Inter-Organization Programme for the Sound Management of Chemicals (IOMC)** was established in 1995 by UNEP, ILO, the Food and Agriculture Organization of the United Nations, WHO, the United Nations Industrial Development Organization, the United Nations Institute for Training and Research and the Organisation for Economic Co-operation and Development (Participating Organizations), following recommendations made by the 1992 UN Conference on Environment and Development to strengthen cooperation and increase coordination in the field of chemical safety. The purpose of the IOMC is to promote coordination of the policies and activities pursued by the Participating Organizations, jointly or separately, to achieve the sound management of chemicals in relation to human health and the environment.

WHO Library Cataloguing-in-Publication Data

Principles for modelling dose-response for the risk assessment of chemicals.

(Environmental health criteria ; 239)

1.Chemicals. 2.Dose-response relationship, Drug. 3.Dose-response relationship, Radiation. 4.Risk assessment. 5.Environmental exposure. I.World Health Organization. II.Inter-Organization Programme for the Sound Management of Chemicals. III.Series.

ISBN 978 92 4 157239 2 (NLM classification: QV 38)
ISSN 0250-863X

This document was technically and linguistically edited by Marla Sheffer, Ottawa, Canada, and printed by Wissenchaftliche Verlagsgesellschaft mbH, Stuttgart, Germany.

CONTENTS

ENVIRONMENTAL HEALTH CRITERIA ON PRINCIPLES FOR MODELLING DOSE–RESPONSE FOR THE RISK ASSESSMENT OF CHEMICALS

NOTE TO READERS OF THE CRITERIA MONOGRAPHS

Every effort has been made to present information in the criteria monographs as accurately as possible without unduly delaying their publication. In the interest of all users of the Environmental Health Criteria monographs, readers are requested to communicate any errors that may have occurred to the Director of the International Programme on Chemical Safety, World Health Organization, Geneva, Switzerland, in order that they may be included in corrigenda.

Environmental Health Criteria

PREAMBLE

Objectives

In 1973, the WHO Environmental Health Criteria Programme was initiated with the following 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.

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

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

The original impetus for the Programme came from World Health Assembly resolutions and the recommendations of the 1972 UN Conference on the Human Environment. Subsequently, the work became an integral part of the International Programme on Chemical Safety (IPCS), a cooperative programme of WHO, ILO, and UNEP. In this manner, with the strong support of the new

partners, the importance of occupational health and environmental effects was fully recognized. The EHC monographs have become widely established, used, and recognized throughout the world.

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

Scope

Two different types of EHC documents are available: 1) on specific chemicals or groups of related chemicals; and 2) on risk assessment methodologies. The criteria monographs are intended to provide critical reviews on the effect on human health and the environment of chemicals and of combinations of chemicals and physical and biological agents and risk assessment methodologies. As such, they include and review studies that are of direct relevance for evaluations. However, they do not describe *every* study carried out. Worldwide data are used and are quoted from original studies, not from abstracts or reviews. Both published and unpublished reports are considered, and it is incumbent on the authors to assess all the articles cited in the references. Preference is always given to published data. Unpublished data are used only when relevant published data are absent or when they are pivotal to the risk assessment. A detailed policy statement is available that describes the procedures used for unpublished proprietary data so that this information can be used in the evaluation without compromising its confidential nature (WHO (1990) Revised Guidelines for the Preparation of Environmental Health Criteria Monographs. PCS/90.69, Geneva, World Health Organization).

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

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

Procedures

The following procedures were followed in the development and publication of this EHC. A designated IPCS Staff Member (Dr Sam Page and subsequently Dr A. Tritscher), responsible for the scientific content of the document, served as the Responsible Officer (RO). The IPCS editor was responsible for layout and language.

The WHO Planning Group for the IPCS Harmonization Project on Dose–Response Modelling met on 10 October 2002 in Geneva to develop an outline and proposed time frame for the project. A first draft working paper, including contributions from several additional authors, was prepared by Drs C. Carrington and M. Bolger and distributed to the Task Group prior to the Task Group meeting, which was held from 13 to 17 September 2004. The first draft working paper was revised during the Task Group meeting and during a subsequent internal Task Group Internet forum. This revised draft was available on the IPCS web site for external review and comment. Comments received are available on request from the WHO Secretariat. They were reviewed by the Task Group, and necessary additions and revisions to the document were made.

The Task Group members serve as individual scientists, not as representatives of any organization, government, or industry. All individuals who as authors, consultants, or advisers participate in the preparation of the EHC monograph must, in addition to serving in their personal capacity as scientists, inform the WHO Secretariat if at any time a conflict of interest, whether actual or potential, could be perceived in their work. They are required to sign a declaration of interest statement. The Chairpersons of Task Groups are briefed on their role and responsibility in ensuring that these rules are followed. Such a procedure ensures the transparency and probity of the process. Their function is to evaluate the accuracy, significance, and relevance of the information in the document. A

summary and recommendations for further research and improved safety aspects are also required. The composition of the Task Group is dictated by the range of expertise required for the subject of the meeting and, where possible, by the need for a balanced geographical distribution.

WHO PLANNING GROUP FOR THE IPCS HARMONIZATION PROJECT ON DOSE–RESPONSE MODELLING

Members

Dr P.M. Bolger, Food and Drug Administration, College Park, MD, United States of America (USA)

Professor E. Dybing, Norwegian Institute of Public Health, Oslo, Norway

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

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

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

Dr C. Portier, National Institute of Environmental Health Sciences, Research Triangle Park, NC, USA

Professor A. Renwick, University of Southampton, Southampton, United Kingdom

Secretariat

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

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

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

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

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

Dr S. Page and Dr A. Tritscher, IPCS, served as the Responsible Officer (RO) and were responsible for the preparation of the final document and for its overall scientific content. Marla Sheffer, Ottawa, Canada, was the IPCS editor responsible for layout and language.

* * *

Risk assessment activities of IPCS are supported financially by the Department of Health and Department for Environment, Food & Rural Affairs, Food Standards Agency, United Kingdom; Environmental Protection Agency, Food and Drug Administration, and National Institute of Environmental Health Sciences, USA; European Commission; German Federal Ministry of Environment, Nature Conservation and Nuclear Safety; Health Canada; Japanese Ministry of Health, Labour and Welfare; and Swiss Agency for Environment, Forests and Landscape.

* * *

Task Group members

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

Professor A. Boobis, Imperial College London, London, United Kingdom

Dr C. Carrington, Food and Drug Administration, College Park, MD, USA

Dr V. Cogliano, International Agency for Research on Cancer, Lyon, France

Professor E. Dybing, Norwegian Institute of Public Health, Oslo, Norway

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

Professor E. Faustman, University of Washington, Seattle, WA, USA

Dr M. Healy, Food Standards Australia New Zealand, Canberra, Australia

Mr J. Howlett, Food and Agriculture Organization of the United Nations, Rome, Italy

Dr J. Kleinberg, European Food Safety Authority, Brussels, Belgium

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

Dr J. Larsen, Danish Institute of Food and Veterinary Research, Solberg, Denmark

Dr D. Lovell, University of Surrey, Surrey, United Kingdom

Dr C. Portier, National Institute of Environmental Health Sciences, Research Triangle Park, NC, USA

Professor A. Renwick, University of Southampton, Southampton, United Kingdom

Professor T. Sanner, Institute for Cancer Research, Oslo, Norway

Dr J. Schlatter, Swiss Federal Office of Public Health, Zurich, Switzerland

Dr R. Setzer, Jr, Environmental Protection Agency, Research Triangle Park, NC, USA

Professor W. Slob, National Institute of Public Health and the Environment (RIVM), Bilthoven, Netherlands

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

Professor G. Williams, New York Medical College, Valhalla, NY, USA

Secretariat

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

Dr A. Tritscher, International Programme on Chemical Safety, World Health Organization, Geneva, Switzerland

Other contributors

Professor E. Calabrese, University of Massachusetts, Amherst, MA, USA

Dr I. Dewhurst, Pesticides Safety Directorate, York, United Kingdom

Dr D. Gaylor, Gaylor and Associates, Eureka Springs, AR, USA

Professor P. Grandjean, Environmental Medicine, University of Southern Denmark, Odense, Denmark

Dr I. Mangelsdorf, Fraunhofer Institute of Toxicology and Experimental Medicine, Hanover, Germany

Dr S. Sand, Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden

Dr K. Victorin, Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden

PREFACE

The International Programme on Chemical Safety (IPCS) was initiated in 1980 as a collaborative programme of the United Nations Environment Programme (UNEP), the International Labour Organization (ILO), and the World Health Organization (WHO). One of the major objectives of IPCS is to improve scientific methodologies for assessing the effects of chemicals on human health and the environment. As part of this effort, IPCS publishes a series of monographs, called Environmental Health Criteria (EHC) documents, that evaluate the scientific principles underlying methodologies and strategies to assess risks from exposure to chemicals.

This EHC is part of the ongoing review of the underlying scientific bases for decision-making in chemical risk assessment by IPCS. It involves specific consideration of the area of dose–response assessment in the evaluation of information from toxicological studies in animals and from human clinical and epidemiological studies. It covers toxicants with threshold effects and those for which there may be no practical threshold, such as substances that are genotoxic and carcinogenic. The discussions are concerned with that subset of cause–effect relationships commonly referred to as dose–response models, which are typically used to characterize the biological effects of intentional (e.g. drugs and nutrients) and unintentional (e.g. contaminants) exposure to chemicals.

This EHC is intended primarily to provide descriptive guidance for risk assessors in using dose–response modelling in hazard characterization. It will also provide mathematical modellers with an appreciation of issues to be considered when 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 dose–response modelling. For both risk assessors and risk managers, some considerations for communicating the results of risk assessments that use dose–response modelling are presented.

The efforts of all who helped in the preparation, review, and finalization of the monograph are gratefully acknowledged. Special thanks are due to Health Canada, the Ministry of Health of Japan, the United Kingdom Food Standards Agency and the United States

National Institute of Environmental Health Sciences for their financial support of the project.

ACRONYMS AND ABBREVIATIONS

ADI	acceptable daily intake
AIC	Akaike's information criterion
ALT	alanine aminotransferase
BMD	benchmark dose
BMD_{10}	benchmark dose at 10% risk
BMDL	lower confidence limit on the benchmark dose
BMDS	Benchmark Dose Software (United States Environmental Protection Agency)
BMR	benchmark response
CDF	cumulative distribution function
CSAF	chemical-specific adjustment factor
DDT	dichlorodiphenyltrichloroethane
DNA	deoxyribonucleic acid
DRM	dose–response modelling
EHC	Environmental Health Criteria
ED_{10}	effective dose for a 10% risk
EPI	exposure potency index
F	Frequency
FAO	Food and Agriculture Organization of the United Nations
f_x	dose–response function
f(x)	dose–response function
IPCS	International Programme on Chemical Safety
JECFA	Joint FAO/WHO Expert Committee on Food Additives
LD_{50}	median lethal dose
LED_{10}	lower bound on the effective dose resulting in a 10% increase in risk (ED_{10})
LL	log-likelihood
LOAEL	lowest-observed-adverse-effect level
MOE	margin of exposure
NOAEL	no-observed-adverse-effect level
NOEL	no-observed-effect level
RfD	reference dose
SD	standard deviation

T_{25}	chronic daily dose that gives 25% of the animals tumours above background at a specific tissue site
TDI	tolerable daily intake
UF	uncertainty factor
WHO	World Health Organization

1. SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

1.1 Summary

Dose–response modelling (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—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 an assessment of the quality of the dose–response analysis and the sensitivity of model predictions to the assumptions used in the analysis.

The characterization of dose–response relationships in animal and human studies has been a major component of hazard characterization and has been used in the extrapolation of incidences of adverse effects in the range of human exposure levels. Over the years, a variety of methods have been developed to accommodate such relationships, for improving extrapolation to low doses and deriving health-based guidance values, such as acceptable daily intakes (ADIs), tolerable daily intakes (TDIs), and reference doses (RfDs). DRM may prove useful in risk assessments for making better use of available data and for providing tools to evaluate the quality of data and the ensuing uncertainties in dose–response estimates.

In general, DRM estimates are based on data from the entire dose–response curve for the critical effect. The standard no-observed-adverse-effect level (NOAEL) approach can be regarded as a special, simplified case of dose–response analysis, as it identifies a single dose that is assumed to be without an appreciable

adverse effect. The dose–response model reflects the characteristics of the dose–response curve, particularly in providing estimates of the slope. In the case of a regression framework, it provides standard error and confidence intervals for the model parameters. A disadvantage of using the NOAEL approach is that it is not possible to quantify the degree of variability and uncertainty that may be present, whereas other dose–response models can facilitate the analysis of sensitivity and uncertainty. Consideration of a dose–response model can optimize study design and clarify the need for additional studies. The NOAEL approach incorporates biological information through the application of "expert" but subjective judgement. Full DRM has the potential for a more "science-rich" analysis through the more formal quantitative inclusion of, for example, factors and covariates into the models. Estimates derived from DRM enhance the ability to compare quantitatively different experiments, effects, and compounds within a common framework. DRM can enhance risk and safety assessments as well as provide opportunities to consider the likelihood of effects outside the observable range.

The choice of the models to be used depends upon the type of data. The models should include a model for dose–response 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 can be evaluated using goodness-of-fit measures. In addition, their ability to describe the data with respect to each other may be compared. Uncertainties about the inferences that result from such models fall into four main categories: statistical uncertainty of inferences due to variability among the responses of experimental subjects, experimental errors (e.g. imperfect randomization, dosing errors, unfavourable dose location), variability among experiments due to unavoidable differences in experimental execution, and uncertainty due to the fact that the "true model" for the data is unknown. Dose–response analysis needs to address all four sources of variability and uncertainty whenever possible.

One particularly important application of DRM is the calculation of benchmark doses (BMDs). These are doses at which it is inferred that a particular level of response would occur. When appropriate data are available, BMDs are an alternative to the NOAEL approach in the calculation of health-based guidance values. When extrapolation is necessary, the uncertainty associated

with a prediction should be represented. Here it is especially important to include model uncertainty.

Full DRM offers the potential to provide additional information for the risk manager. The output of DRM should be directed towards addressing specific questions about the likelihood of adverse health effects. It can be presented in three principal ways. Firstly, it can be used for the establishment of a health-based guidance value, such as an ADI, TDI, or RfD, analogous to current procedures based on a NOAEL or lowest-observed-adverse-effect level (LOAEL). DRM can be a more scientifically robust method for determining health-based guidance values. Secondly, the output from DRM can be used in risk management to estimate a margin of exposure (MOE), by calculation of the ratio of the dose corresponding to a given limit of response to a human exposure level. Thirdly, based on the modelled dose–response relationship, the output can be a quantitative estimation of the magnitude of the risk/health effect at the level of human exposure, with the generally accepted assumption that the uncertainty factors used cover the uncertainties about differences in sensitivity between individuals and species. DRM can provide better information on the likelihood of effects at low doses that are below the levels observed in biological systems and can also provide better estimates of the statistical uncertainties surrounding estimates of likely effects.

Two factors that can impact the type of outputs from DRM exercises and that may be of importance to the risk manager are multiple data sets and uncertainties. DRM can be used with exposure data to identify subpopulations at risk. DRM can also be used to assist risk managers in determining priorities and evaluating the consequences of proposed interventions aimed at reducing the risk. For risk communication, the use of DRM techniques offers opportunities and challenges. DRM evaluations can produce information in several formats, including dose–response functions that allow, along with estimates of exposure, the prediction of risks at specified exposure levels and functions that allow the estimation of exposure levels resulting in specified risks. This includes estimates of the possible risk at intakes above a health-based guidance value, such as an ADI. DRM evaluations also offer approaches to compare competing risks or benefits and provide a focus on uncertainties that can influence the predicted risk.

However, unless the situation of risk is viewed at the population level, there is the risk communication problem that while explaining the level of risk in those circumstances where there is no safe level of exposure, some percentage of the population will be predicted to experience some effect deemed to be adverse. It must be recognized that the use of DRM requires a certain quantity and quality of data, as well as specific expertise.

The potential "ongoing" use of the estimates from DRM can, from a risk management perspective, give an improved characterization for decision-making by:

- providing information about what happens above the health-based guidance value (magnitude and types of health impacts);
- showing benefits from different regulatory actions;
- giving the decision-maker a "more-than-one-point" appreciation of the data;
- promoting consistency in decisions, if appropriate adjustments are made for differences in effect, effect level, species, and study design; and
- allowing for an iterative interaction between the risk assessor and risk manager on a continuous and ongoing basis.

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

- explaining that some percentage of the population is predicted to exceed the safety level and/or experience an adverse effect;
- explaining the level of risk in those circumstances where there is assumed to be no safe level of exposure;
- comparing competing risks or benefits;
- providing a focus on uncertainties that influence the predicted risk; and
- explaining that a risk estimate pertains to what may occur at a population level, rather than the individual level, and noting that this is also the case for the ADI/TDI approach.

1.2 Conclusions

- Full DRM can be considered a more sophisticated or robust alternative to the NOAEL approach in all cases where suitable dose–response data are available (e.g. several dose groups with different response levels).

- For quantal dose–response data, the interest is often in low response (incidence) levels. This may call for low-dose extrapolation by several orders of magnitude (e.g. for tumour incidences). However, equally plausible dose–response risk models may result in highly divergent low estimates. A currently applied approach is to estimate a BMD_{10} (dose at 10% risk) and linearly extrapolate from that point downwards, as a conservative approach. Another option, currently under development, is to apply a Bayesian approach that considers the various models all together.

- For continuous dose–response data, two approaches of DRM exist. One is to transform the continuous data into quantal data. The other is to consider continuous dose–response data as information on the severity of the effect and therefore as a function of dose. In the latter approach, measurable changes of effect are often close to response levels considered as adverse (e.g. 10% inhibition of cholinesterase), and the low-dose extrapolation problem is minor or non-existent.

- For the purpose of deriving an ADI, TDI, or RfD, DRM may be used for deriving a BMD, to be used as a point of departure in the same way as the NOAEL is used (i.e. the same uncertainty factors would be applied to the BMD as to the NOAEL).

- DRM may also be used for estimating risks at a given (human) exposure level. For risks in terms of incidences (quantal data), this may involve low-dose extrapolation.

- DRM exercises can provide information on uncertainties associated with the data and identify factors contributing to uncertainties in risk estimates.

- Application of DRM for all end-points can be cost prohibitive, so it is efficient to pre-select the apparently more sensitive end-points. In some cases, however, it is not easy to identify the most sensitive end-points by visual inspection, so all of the end-points may need to be modelled.

- The BMD and the lower confidence limit of the BMD (BMDL) should always be reported, so that the quality of the data and the model fit are clear and potencies can be compared on the basis of the BMD.

- The output of the different models used in DRM should be presented.

1.3 Recommendations

- Toxicity testing protocols (e.g. Organisation for Economic Co-operation and Development guidelines) should be reviewed for optimization for BMD and other DRM approaches, including optimal designs for the number of animals and number of doses for different dose–response curves. Additional research is needed for the development of optimal study designs. Guidance should be developed for combining existing studies with a view to DRM.

- Better guidance needs to be developed for combined analysis of different data sets for more precisely estimating BMDs.

- Better understanding of when and how to use the benchmark response (BMR) needs to be developed.

- Better understanding of the shape of the dose–response curve at low doses needs to be developed. Additional research is needed to determine the biological basis for extrapolation (e.g. by using biomarkers, tumour precursors, genetically modified animals, and toxicokinetics for target dose estimation).

- Improved guidance needs to be developed for risk communication based on the results of DRM and probabilistic assessment techniques. This should include communication of

the types of uncertainty and the relation to statistical variability, imprecision, and the use of confidence intervals.

- The use of DRM should be reviewed and additional general principles for its use developed when more experience becomes available.

2. 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 information is needed and will be subject to an analysis of a dose–response relationship.

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, in the absence of human data, typically evaluated in laboratory animals at significantly higher doses than the levels to 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 quantification of the risk above the level of exposure that has been assessed to be safe (e.g. 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. Although dose–response models have been available for some time, their use has been somewhat limited because of a lack of either appropriate scientific information or agreed-upon approaches and methods for how to obtain and use available dose–response 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).

Dose–response 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; like any other tool used in science, DRM needs to be utilized in a broader context of all of the available scientific knowledge. Although mathematical and statistical rigor are important factors in risk assessment, the final standard that prevails remains biological plausibility. It is this inherent uncertainty and its communication for which modelling and quantitative risk assessment can be particularly valuable.

2.1 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 the issues to be considered when 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.

2.2 Scope

This EHC is part of the ongoing review of the underlying scientific bases for decision-making in chemical risk assessment by IPCS. It involves specific consideration of the area of dose–response assessment in the evaluation of information from toxicological studies in animals and from human clinical and epidemiological studies; it does not include consideration of other aspects of quantitative risk assessment, such as physiologically based modelling. It covers toxicants with threshold effects and those for which there may be no practical threshold, such as substances

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

This document focuses primarily on experimental animal studies. In DRM of human epidemiological data, several important issues should be considered:

- *Impact of imprecision of the dose estimate*. This issue differs substantially from the situation with experimental animal studies. In observational studies, where the dose is not a matter of design, this imprecision is likely to be substantial.

- *Absence of a true control group*. In many observational studies, there may not be any subjects who are completely free from exposure. The response at zero exposure cannot be observed and has to be estimated.

- *Shape of the dose–response curve at low doses*. The shape may depend on both the outcome parameter and the toxicant. For most contaminants, insufficient information is available, and the impact on uncertainties must therefore be considered.

- *Confounder adjustment*. In epidemiological studies, confounder adjustment must be included. Decisions therefore need to be made as to which confounders to include in the DRM.

- *Meta-analysis*. If more than one study is available, a meta-analysis can provide improved information on the dose–response models.

Many of the considerations in this EHC are also relevant to ecotoxicological studies.

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 3) and the basic concepts of DRM (chapter 4). In

chapter 5, the use of DRM is described, including comparing the no-observed-adverse-effect level (NOAEL) approach with the benchmark dose (BMD) modelling method. Chapter 6 provides the principles of DRM, including data considerations, model descriptions, model fitting and parameter estimation, model comparisons, and uncertainty. This chapter also includes discussion of BMD approaches. Chapter 7 discusses the provision of scientific advice by risk assessors to the risk managers. This chapter includes an explanation of the output of the dose–response analysis and the strengths and weaknesses of DRM. The final chapter, chapter 8, summarizes the conclusions of the EHC and provides recommendations for future research.

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.

3. RISK ANALYSIS

3.1 Decision paradigms

A risk analysis decision paradigm is a formal representation of a process that distinguishes the scientific bases from the risk management objectives and generally contains a component in which 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).

3.2 Risk analysis paradigms

The first risk analysis paradigm for public health was proposed by the National Academy of Sciences of the United States of America (NRC, 1983) and focused on assessing the risk of cancer from exposure to chemicals in food. The decision process was divided into three major steps: 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 National Academy of Sciences 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 that 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 (NRC, 1994, 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 1).

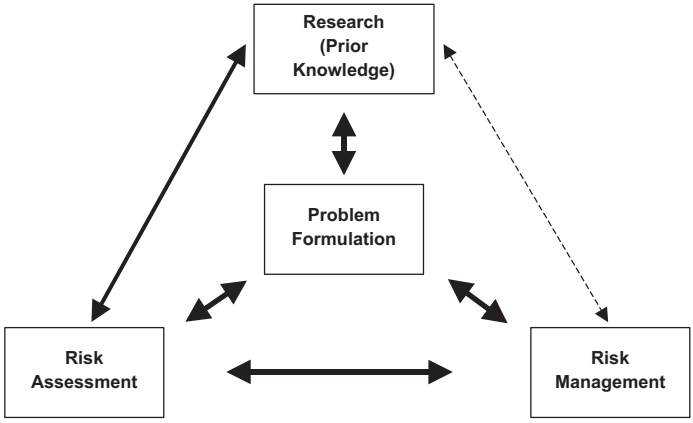


Fig. 1. Interactions of risk assessment with risk management.

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.

As the risk analysis paradigm evolved, the need for risk communication as an integral part was recognized (see Figure 2). Risk communication not only is 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 as the basis of the risk management decisions demands transparency and appreciation for the uncertainties involved.

Fig. 2. The risk analysis paradigm.

3.3 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.

3.3.1 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 are 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 assessments. 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 its level of interest increases and sophistication decreases. Producing records of an assessment with varying degrees of technical detail may be a useful objective.

The World Trade Organization, under the Agreement on the Application of Sanitary and Phytosanitary Measures, has recognized the importance of harmonized science-based risk assessments. The World Trade Organization has specifically cited the standards, guidelines, and recommendations of the Codex Alimentarius Commission as reflecting international consensus regarding the requirements to protect human health from foodborne hazards. The Codex Alimentarius Commission has formally adopted the risk analysis paradigm in its decision-making (Codex Alimentarius Commission, 2003). Other organizations have also adopted this paradigm (European Commission, 2000).

3.3.2 Public health and individual health

A public health risk assessment is concerned with a population. The behavioural, environmental, or biological characteristics will vary among individuals in the population of concern. This variation is considered in probabilistic approaches and determines the statistical nature of health risk measures and conclusions made on populations. A risk assessment may need to describe or model these individual characteristics to produce a prediction of what might 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.

3.3.3 Quantification and computation

Public health issues often involve matters of degree, particularly in regard to level of exposure and risk, and may be defined by measures of quantity or statistical rates. If an uncertainty analysis is conducted, knowledge may be quantified as a matter of degree. Although judging matters of degree does not require the use of numbers, communication of degree does. Quantitative risk assessment approaches, including DRM, can be valuable in providing information to address these issues.

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

3.3.4 Cost of assessment

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

3.4 Risk assessment

The risk assessment paradigm, incorporating problem formulation, is illustrated in Figure 3 (based on Renwick et al., 2003).

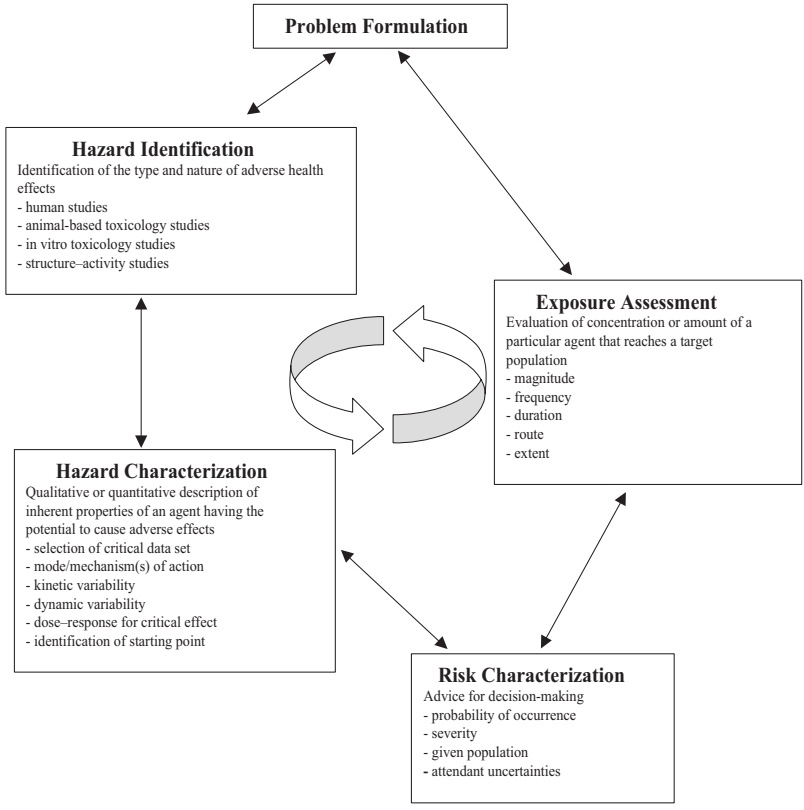


Fig. 3. Risk assessment (adapted from Renwick et al., 2003).

3.4.1 Problem formulation

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

3.4.1.1 Defining the question

Among additional considerations are those that address who should be involved in the risk assessment and risk management processes. The transparency of a risk assessment will depend on how well these are described. It is not necessary to establish beyond all doubt that there is a cause–effect relationship in order to conduct a risk assessment. The suspicion that there may be such a relationship is sufficient. The consideration of the evidence for or against the supposition is often an integral part of the analysis. Identifying the problem may be politically controversial. That is, it may constitute a risk management issue that must be resolved before the risk assessment may be used as the justification for a decision. Non-scientific controversy may be diverted from the risk assessment by separating the valuation of the effect from the risk assessment per se (i.e. the risk assessment may be used as part of a cost–benefit analysis, but the cost–benefit analysis is not part of the risk assessment). Predicting the occurrence of an event is not part of an expression of the level of public health concern. However, suggesting that the problem is big enough to merit a formal risk assessment does imply that the risk may be of some significance.

3.4.1.2 Prior knowledge

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

3.4.1.3 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 the assessment.

3.4.2 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 (e.g. threshold of toxicological concern); and
- risks at minimum and maximum intakes (e.g. nutrients).

Qualitative advice includes:

- statements/evidence that the agent is of no toxicological concern owing 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 a 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 subpopulations, including those with greater potential exposure and/or specific physiological conditions or genetic factors. At present, this is limited, and generic approaches have to be used (e.g. 10 × 10 uncertainty factors for interspecies differences and human variability). The advice to risk managers can be in the form of a comparison of the relative risks resulting from choosing different 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 conclude 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. It is also possible that additional information can increase uncertainty.

4. DOSE–RESPONSE MODELLING: BASIC CONCEPTS

4.1 Introduction

Toxicology is the science of identifying and quantifying harmful or adverse effects of chemical and physical agents in the human environment. This can be accomplished by observations in humans (i.e. epidemiological and clinical studies), experimental studies using animal models (i.e. in vivo bioassays), or cellular and molecular studies. All these approaches have firmly established the principle of dose–response. Accordingly, dose–response toxicities of chemicals can be and have been expressed quantitatively (e.g. the median lethal dose, or LD_{50}).

However, scientific data alone are not sufficient to allow a decision to be made regarding the potential toxicity of chemicals and other agents that humans encounter; it is the analysis and interpretation of these data that lead to a scientifically supported decision regarding potential health effects. Many analytical processes have been developed to address the evaluation of the toxicities of chemicals, ranging from very simple approaches based solely upon the identification of the possibility of a hazard (NTP, 2002; Cogliano et al., 2004; USEPA, 2005) to much more complicated approaches incorporating biological mechanisms, complicated mathematical models, bioavailability in humans, and direct predictions of chemically induced changes in disease incidence in the affected human population (Portier & Kohn, 1996; Kim et al., 2002). All of these methods have two basic steps in common: analysis of the dose–response information and implementation of the results of that analysis to formulate a conclusion. The combined two-step approach will be referred to as DRM.

This chapter describes the elements that embody DRM. Most of the information presented is found in more extensive detail in other chapters of this guidance document. This chapter sets the stage for discussion of dose/exposure–response modelling by briefly

answering the questions: What is dose? What is response? What is a model? It then goes on to introduce the reader to the types of data and information that may have an impact on the development of dose–response models.

4.2 What is dose?

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

Two other parameters are important: the dose frequency and duration of dosing. Dosing can be acute, subchronic, or chronic. For simplicity, the term dose in DRM will be used as an inclusive term referring to all three forms of dose described above. In general, units of dose should reflect the magnitude, frequency, and duration over which it applies. Dose can be expressed in a multitude of metrics. Some of these metrics include daily intake (e.g. ng/kg body weight per day), total body burden (e.g. ng/kg body weight), body burden averaged over a given period of time, or tissue concentration (ng/kg).

For humans, where dosing of xenobiotics is not intentional, the term exposure is used for the external dose. In epidemiological studies, exposure is rarely known, and best estimates are made using several assumptions and/or biomonitoring of tissue (usually blood) concentrations at very few time points, often many years

after what is believed to be the period of first/highest exposure. Sometimes, when laboratory animals are used for DRM, the dose used in the animal study is transformed to an equivalent human exposure prior to modelling. Exposure assessment is the qualitative and/or quantitative evaluation of the likely intake of chemical agents via food, as well as exposure from other sources, if relevant (WHO, 1997). In this situation, models of exposure linked to response data may be used to develop a dose–response model. However, limited knowledge of the events controlling absorption and tissue distribution (especially in humans at low levels of exposure), metabolism, and excretion and the other molecular and biochemical processes that ultimately lead to particular responses contribute to the uncertainty in these analyses.

4.3 What is response?

Response, in this context, generally relates to an observation or effect seen in a laboratory cell culture, an animal, or a human following exposure. These end-points cover a broad range of observations, from early responses, such as biochemical alterations, to more complicated responses, such as cancer and developmental defects. Responses can be either adaptive or adverse (e.g. Williams & Iatropoulos, 2002). The latter are defined as a change in the morphology, physiology, growth, development, reproduction, or lifespan of an organism or subsystem (e.g. subpopulation of cells) 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. These are critical responses that are likely to underlie an adverse health effect in humans. The responses are sometimes species and/or tissue specific and have different degrees of variation across individuals. Nevertheless, there is some commonality across species, and there are known linkages between some responses (e.g. DNA damage is a precursor for mutations). DRM can address each response, provide insight into their quantitative similarity across species and tissues, and link responses in a mechanistically reasonable manner.

Response is generally considered to vary across experimental units (animals, humans, cell cultures) in the same dose group in a random fashion. This random variation is usually assumed to follow some statistical distribution describing the frequency of any given

response for a population. In general, statistical distributions are characterized by their central tendency (usually the mean or average value) and their effective range (usually based on the standard deviation). Most responses of interest in the context of dose–response assessment fall into one of four basic categories:

- *Quantal responses.* Quantal responses generally relate to the number of experimental units responding in a given period of time (e.g. the proportion of animals with a tumour in a cancer bioassay).

- *Counts.* Count data generally relate to a discrete number of items measured in a single experimental unit (e.g. number of papillomas on the skin).

- *Continuous measures.* Continuous measures generally take on any value in a defined range (e.g. body weight).

- *Ordered categorical measures*. Ordinal categorical measures generally take on one value from a small set of ordered values (e.g. tumour severity grades).

Sometimes it is useful to convert continuous data into proportions (e.g. number of animals outside a clinically relevant range for an immune system marker) or categories (e.g. measured degree of liver necrosis converted to minimal, moderate, or extensive).

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

4.4 What is a model?

Dose–response models are mathematical models used to characterize the relationship between dose and response for a given set of scientific data. Mathematical models consist of three basic components: assumptions used to derive the model, a functional form for the model, and parameters that are components of the

functional form. For example, the simplest dose–response model is a linear model to describe a continuous response (see Figure 4).

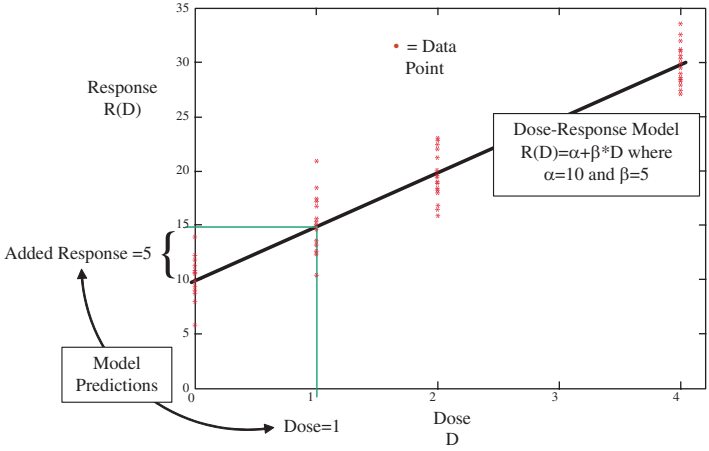


Fig. 4. Dose–response illustration displaying a linear model fit to continuous data for which prediction of the dose associated with an added response of 5 units (not designated) is a dose of 1 unit (not designated).

For this model, the key components are:

- *Assumptions*: Mean added response is proportional to dose.
- *Functional form*: $R(D) = \alpha + \beta \cdot D$, where R(D) is the mean response as a function of dose, denoted D.
- *Parameters*: α is a parameter describing the mean response in the control (unexposed) group, and β is a parameter describing the mean change in response per unit dose.

Dose–response models range from very simple models, such as the linear model described above, to extremely complicated models for which the eventual functional form cannot easily be expressed as a single equation (e.g. biologically based dose–response models). Models can also be linked, meaning that one model could describe part of the dose–response process while another describes the remainder of the process. For example, in most cases for chemical carcinogenesis, cancer risk is more closely linked to tissue

concentration than to administered dose. Given data on dose, tissue concentration, and tumour response, one can use a toxicokinetic model to relate dose to tissue concentration and use a multistage cancer model to relate tissue concentration to response. The two models combined are needed to describe the dose–response relationship.

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

4.5 What is dose–response modelling?

DRM can be described by six basic steps, with a variety of options at each step (Table 1). The first four steps, which will be referred to as dose–response analysis, are aimed at the analysis of the data available for DRM. Dose–response analysis provides the linkage of a model to dose–response data for the purposes of predicting response to a given dose or predicting dose from a given response. The last two steps deal with implementation and evaluation of the analysis results.

Table 1. Basic steps in dose–response modelling

Step	Description	Options	Section links for chapter 6
1. Data selection	Determine the response to be modelled, and select appropriate data	End-point, data quality, sample size, data utility, data availability	6.1
2. Model selection	Choose the type of model to be applied to the data	End-point, data availability, purpose	6.2.1
3. Statistical linkage	Assumes that statistical distributions describe the response	End-point, data type, model choice, software availability	6.2.2
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	6.3
5. Implementation	Use the estimated model parameters and the model	Outputs, target selection, model predictions, BMD,	6.3

Step	Description	Options	Section links for chapter 6
	formula to predict response/dose as needed	direct extrapolation	
6. Evaluation	Examine the sensitivity of the resulting predictions to the assumptions used in the analysis	Model comparison, uncertainty	6.4, 6.5, 6.6

Step 1 involves selection of the appropriate data for DRM. The type of data available can have a marked impact on the complexity of the model that can be used. For example, whereas two points can be used to identify the slope of a line, it takes at least three points to identify the shape of a more complex dose–response relationship (e.g. straight line versus two connected lines). The issue of whether there are enough data to support a given model is quite complex (Portier, 1994) and is discussed in greater detail in section 6.1. In general, the data can restrict the type of model that can be used.

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

The third step requires the choice of a statistical linkage between the data and the model. The most common linkage method is to assume a statistical distribution for the response and use that distribution to derive a mathematical function describing the quality of the fit of the model to the data. However, a considerable amount of DRM has been done by simpler linkage functions, such as drawing a straight line through the data points. The advantage of

choosing a formal statistical linkage is the ability to test hypotheses and derive confidence intervals for model predictions.

In DRM, fitting the model to the data is the fourth step. Since the primary components of a model are the parameters that define the model, curve fitting simply involves choosing values for the parameters in the model. If a formal statistical linkage has been developed for linking the data to the model, then the parameters are chosen such that they "optimize" the value of the linkage function. For example, a common choice is to link the data to the model using the squared distance, denoted $[R(d_i) \bullet o_{ij}]^2$, between the predicted value from the model, denoted $R(d_i)$, and the observed value, denoted o_{ij}. These squared differences can be summed across all data points, and model parameters are chosen to minimize this sum; this is the common least-squares algorithm. Simpler methods can also be used to estimate model parameters. For example, by drawing a line through the data points, the parameters in the linear model can be estimated directly, since the value of can be estimated as the point where the line crosses the y-axis (zero dose) and the value of can be estimated by calculating the slope of the drawn line. Formal optimization is a better choice for modelling than ad hoc procedures, which often do not meet the criterion of transparency.

The fifth step in DRM is to make the inferences necessary to develop measures to protect public health. In its simplest form, a dose–response model allows the prediction of the response if the dose is known and the calculation of the dose if the aim is to target a specific level of response. In addition, implementation of the dose–response analysis (steps 1–4) also encompasses the extrapolation of results from the specific responses seen for the experiment being modelled to other exposure scenarios and other doses. This step can also involve an extrapolation from a laboratory species to humans. Usually, when making a prediction, the emphasis is on the change in response seen in the treated animals compared with the response seen in the controls. The different types of data (quantal, count, continuous, categorical) require different methods for predicting changes in response beyond the normal response. In general, the targets used for additional response fall into the categories of added response (simply subtract control

response), relative response (fold change relative to control response), and extra response (added response scaled to range from zero to the maximum possible response). Each of these choices can impact the final decision, so care should be taken to understand why a specific choice is made. Figure 4 illustrates some of the basic components of DRM for the simple linear model case and added response.

Measures used by public health agencies to prevent excess exposure to a hazardous agent generally fall into the categories of direct banning or limiting exposure. DRM could inform both choices, although its major impact is in the area of limiting exposure. Several methods on how to use DRM in this context have been proposed. The simplest is to use the predicted model to find the dose associated with a negligible (e.g. one in a million) or zero response over control. In general, this results in extrapolation far beyond the range of the data, which creates a great deal of uncertainty. A second approach is to use the dose–response model to identify a dose with a known response at or slightly below the observable range (the limit of scientific certainty) and use other models to get into a range where the response is assumed to be virtually unchanged relative to the control response. In this approach, a functional model structure can be used, such as a straight line, or something simpler, such as uncertainty factors (UFs), to identify a safe level of exposure. All of these options are discussed in chapter 6.

The basic steps in DRM shown in Table 1 can be repeated to consider other options in the process in order to understand the impact of choices on the predictions from DRM. This final step (step 6) in DRM is aimed at understanding the sensitivity of the analysis to specific choices and judging the overall quality of the final predictions. The simplest way to evaluate sensitivity is by considering several choices and determining if the results dramatically change. Depending on the degree of difference between choices, there could be value in performing a formal analysis of the quality of the fit of the model to the data. Other methods can also be used to assess the impact of choices used in the modelling on the eventual outcome, such as uncertainty analysis and Bayesian mixing. In some cases, step 6 is performed before step 5, with a focus on the assumptions used in the dose–response

analysis, and/or after step 5, with a focus on the assumptions used for implementation. These steps are further described in chapter 6.

4.6 Risk versus safety in dose–response modelling

Risk as used in this discussion is the direct estimation of the likelihood or degree of an event or its prevalence in a human population as a function of exposure. Given sufficient data in humans in the range of exposure where there is concern, it is possible to obtain scientifically supported estimates of risk. In most cases, the data used to develop dose–response models are not from studies in humans in the range of exposures that humans generally encounter. The most common type of data used for DRM comes from experiments in laboratory animals, generally at administered doses significantly exceeding the exposures that humans encounter. Even when human data are available and suitable for dose–response analysis, they are generally from selected populations, such as workers in occupational settings, whose exposures differ from those of the general population.

Thus, in many cases, dose–response analyses need to be extrapolated from an observable region where scientific support is available to a region where scientific support is weaker or non-existent. For dose–response analyses based on human studies, the extrapolation is generally a downward extrapolation to different exposure levels, but extrapolations can also be to different life stages (e.g. fetus, child) or to different populations with different environmental factors that might affect exposure (e.g. dietary differences). For dose–response analyses based upon laboratory data using animals, there is the additional problem of extrapolating from animals to humans.

Most of the methods used to implement the results of a dose–response analysis (step 5) address these extrapolation issues. The methods that have been used for extrapolation are diverse and sometimes contentious, with different countries, and even different agencies within a given country, using different approaches. The strategies used for extrapolation basically fall into two categories: those aimed at using estimates of risk for exposures outside of the

range of the data used in the dose–response analysis, and those aimed at establishing safety without using an estimate of risk.

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

Regardless of how dose–response analysis is performed, additional methods are employed to extrapolate to humans. These methods are also varied, ranging from the use of additional uncertainty factors to more complicated modelling schemes based upon differences in toxicokinetics and toxicodynamics between humans and animals.

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

4.7 Summary

DRM, as used for informing public health decisions about chemical exposures, is a six-step process. The first four steps constitute dose–response analysis and 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 implementation of the results of the dose–response analysis for the purposes of guiding a public health decision. The final step, which can optionally be applied earlier in DRM, involves an assessment of the quality of the dose–response analysis and the sensitivity of model predictions to the assumptions used in the analysis. DRM, because it involves a large number of choices based upon scientific experience, can take on many different forms and be used in many different ways. The remaining chapters of this report focus on the range of choices available for each step in the process and some guidance to be used in making these choices.

5. DOSE–RESPONSE MODELLING: WHY AND WHEN TO USE IT

Dose–response analysis 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 dose–response relationships observed in animal bioassays as well as the low-dose extrapolation of incidences of adverse effects to the range of human exposure levels. Dose–response analysis includes the use of the NOAEL (pairwise testing) for deriving health-based guidance values such as the ADI and the use of DRM (fitting functions).

5.1 Historical perspectives

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

The prototype chemical safety assessment uses the ADI methodology, which was introduced by Lehman & Fitzhugh (1954) and has come to be widely employed for the derivation of health-based guidance values (IPCS, 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. The same methodology is used to derive health-based guidance values for chemical contaminants. However, because "acceptable" was deemed to be an inappropriate term for chemical contaminants, the term "tolerable" was used instead (i.e. tolerable daily intake, or TDI). Comparable terms that have been used are provisional maximum tolerable daily intake (IPCS, 1987) and reference dose (RfD) (Barnes & Dourson, 1988). Other similar methods exist for different types of exposures, such as for compounds with accumulating properties—for example, provisional maximum tolerable weekly intake or provisional maximum tolerable monthly intake (IPCS, 1987; WHO, 2002).

5.1.1 The no-observed-adverse-effect level approach to acceptable/tolerable daily intake

Calculation of the ADI based on the NOAEL approach for the case of quantal data is summarized in Table 2.

Table 2. NOAEL-derived ADI for the case of quantal data

Step	NOAEL-derived ADI
1. Data selection	Sufficient sample sizes, at least one dose with "no" effect and one dose with effect. Relevant end-points in a relevant species are important for any approach.
2. Model selection	Statistical method $R(D) = \begin{cases} 0 & \text{if response at dose D is not significantly different from control response} \\ 1 & \text{if response at dose D is significantly different from control response} \end{cases}$
3. Statistical linkage	Pairwise statistical tests between dose groups and control group.
4. Parameter estimation	Assessment of point of departure $NOAEL = D_{NOAEL}$ where $R(D) = 0$ for all $D \leq D_{NOAEL}$ and $R(D) = 1$ for all $D > D_{NOAEL}$ This procedure presupposes that all doses below the NOAEL are non-significant and all doses above the LOAEL are significant. This is often not the case.
5. Implementation	$ADI = \frac{NOAEL}{UFs}$ where UF is uncertainty factor.
6. Evaluation	Statistical power analysis should be performed to check if the test was sensitive enough to detect relevant effects.

Selecting the data needed to calculate the ADI based on the NOAEL approach (step 1) is similar to choosing the data to be used

for more complicated modelling; the better data sets have an appropriate number of relevant doses, sufficient sample sizes, and relevant end-points in a relevant species. The next step in calculating an ADI is to determine the NOAEL, which is the highest concentration or dose of a chemical, found by experiment or observation, that causes no detectable adverse effect, as defined above. This includes a statistical method (step 2), statistical linkage (step 3), and a method of assessment of a point of departure (step 4) that describes the identification of the NOAEL. Consider a response procedure, R(D), of the form:

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

The statistical linkage (step 3) between this procedure and the data is represented by the statistical test used to determine if a response at any given dose is different from the control response. When the response is non-significant, we simply *act as if* the effect were in fact zero. Obviously, we cannot conclude that the effect actually *is* zero. When the NOAEL approach is chosen, the statistical test is used to decide upon the existence of a statistically significant increase (e.g. at the 5% level) over background (e.g. the control group) for each dose level separately. The selection of the NOAEL (step 4) is then achieved by choosing the largest dose, D_{NOAEL}, for which all smaller doses have R(D) = 0 and all larger doses have R(D) = 1. Mathematically, this assessment can be written as:

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

This procedure presupposes that all doses below the NOAEL are non-significant and all doses above the LOAEL are significant. This is not always the case.

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

Historically, an uncertainty factor of 100-fold has been used to convert the NOAEL from an animal study into a health-based guidance value (Lehman & Fitzhugh, 1954; Dourson & Stara, 1983; IPCS, 1987). Additional uncertainty factors may be used to allow for database deficiencies, such as the absence of a chronic study (IPCS, 1994). The default 100-fold uncertainty factor may be seen to represent the product of two separate 10-fold factors that allow for interspecies differences and human variability (IPCS, 1987; Renwick & Lazarus, 1998). The recognition that the original 100-fold uncertainty factor could be considered to represent two 10-fold factors allowed some flexibility, because different factors could be applied to the NOAEL from a study in humans and from a study in animals. The concept of chemical-specific adjustment factors (CSAFs) (IPCS, 1994, 2005) was introduced to allow appropriate data on species differences and/or human variability in either toxicokinetics (fate of the chemical in the body) or toxicodynamics (actions of the chemical on the body) to modify the relevant default 10-fold uncertainty factor. The strategy used by WHO/IPCS in the NOAEL/ADI approach involves replacing the original 100-fold uncertainty factor with CSAFs where there are adequate data (IPCS, 1994, 2005).

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

$$\text{ADI} = \frac{\text{NOAEL}}{\text{UFs}}$$

Step 6 can be extended to the evaluation of the sensitivity of the ADI to the assumed values of the uncertainty factors.

Some scientists have raised concerns regarding the use of the NOAEL to determine an ADI. The greatest concern is that the NOAEL tends to yield lower ADIs for chemicals for which there are more or better data. Therefore, stakeholders using usually more costly, better data are “punished” (Crump, 1984; Dourson et al., 1985; Kimmel & Gaylor, 1988; Barnes et al., 1995; Slob & Pieters, 1998).

5.1.2 The benchmark dose approach to acceptable/tolerable daily intake

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

Calculation of the ADI based on the BMD approach is summarized in Table 3 for the case of quantal data. A generic form of the BMD and benchmark dose lower confidence limit (BMDL) is presented in this table. In this document, a variety of response levels, such as 1%, 5%, and 10%, will be discussed.

Table 3. BMD-derived ADI (Weibull model) for the case of quantal data

Step	BMD-derived ADI
1. Data selection	Sufficient number of doses with different response levels and a sufficient number of *total* subjects.
2. Model selection	Fit dose–response model (e.g. Weibull model).
3. Statistical linkage	Predicted fractions are linked to observed fractions, and their "distance" is minimized by optimizing some fit criteria function (e.g. likelihood function based on assumed distribution).
4. Parameter estimation	Choose an appropriate response, p, in the range of experimental response. Estimate $BMDL_p$, the 95% lower confidence bound on the BMD_p, where $\frac{R(BMD_p) - R(0)}{1 - R(0)} = p$
5. Implementation	$ADI = \frac{BMDL_p}{UFs}$
6. Evaluation	Sensitivity of BMD to model choice can be checked by fitting various models.

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.

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. Complicated models will require a larger number of dose groups than simpler models. Several models have been proposed for each type of data. In the United States Environmental Protection Agency's Benchmark Dose Software (BMDS) program, a number of routinely used models are cited (http://www.epa.gov/ncea/bmds/). 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 α is the proportion responding in the unexposed group, β describes the increase in probability of adverse effect per unit dose, and γ describes the shape of the dose–response curve (e.g. $\gamma >> 1$ implies threshold-like behaviour; $\gamma = 1$ implies log-linear behaviour).

The statistical linkage (step 3) between the data and the model can assume a number of different forms, as described previously (section 4.5) and in section 6.2. For quantal data, it is appropriate to assume that the data are binomially distributed for each dose group. Estimating model parameters (step 4) for the BMD method can also 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 dose–response 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 benchmark response, or BMR (p), below which there was insufficient support from the data. A common choice for BMR is p = 10%. Once the BMR (p) is selected, the BMD, specifically denoted BMD_p, is calculated according to the following equation, if the extra risk formula is used:

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

Empirical investigations showed for a large and representative set of compounds that the 95% statistical lower bound on the estimated BMD may be regarded as an analogue to a NOAEL, and substituting one with the other would result in similar ADIs (Crump, 1984; Barnes et al., 1995). As with all aspects of modelling, many choices exist for calculating confidence bounds, and these are discussed further in chapter 6.

Having chosen a method for estimating a 95% statistical lower bound on BMD_p, which can be called $BMDL_p$, the ADI can be calculated as follows:

$$ADI = \frac{BMDL_p}{UFs}$$

In this calculation, the values of the uncertainty factors could be the same as those used for the NOAEL or adjusted to account for a slightly different interpretation for the $BMDL_p$ relative to the NOAEL (Renwick et al., 2003).

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

5.2 Points of consideration

The use of DRM in general for hazard characterization is possible when a sufficient amount of dose–response information is available, either from an experimental animal bioassay or from a human study (epidemiological study or clinical trial). As shown in the previous section, the BMD can be considered as an alternative point of departure for deriving an ADI in those situations where a NOAEL would have been used as a point of departure in current procedures. In addition, DRM may be helpful in those situations

where there is a need for low-dose extrapolation (e.g. substances that are genotoxic and carcinogenic). It should be noted, however, that extrapolation from a single model that fits the data in the observed range cannot be justified, since other models fitting the data equally well may result in substantially different estimates of low-dose risk. Bayesian approaches are currently under development, which take into account both statistical uncertainties in the data and model uncertainty (see section 6.5). In practice, linear extrapolation from a BMD_{10} (or ED_{10}, effective dose for a 10% risk, approximately equal to BMD_{10}) is often applied as a simple method for low-dose extrapolation. This is considered a conservative approach. As another application, DRM may be used to estimate risks at any given (human) exposure level. Since human exposure levels are usually lower than the doses in the observed range in animal studies, methods for low-dose extrapolation may also be needed in this application.

5.2.1 General aspects of definition

The NOAEL is a parameter derived directly from the observed dose–response data and is defined as the highest administered dose at which the effect is still not significantly different from that at dose 0 (see section 5.1). The NOAEL is based on a multiple test procedure performed along the applied dose series. It lacks further detailed statistical properties compared with a parameter of a dose–response model, for which the precision of the estimate can be quantified.

The dependence of the NOAEL on the statistical significance test, however, tends to penalize chemicals for which there are more or better data by giving a higher estimate for those chemicals with less precise data. This problem does not occur in DRM. In fact, the opposite relationship holds: it penalizes studies with few or poor data.

The NOAEL approach can be formally considered a dichotomous procedure, where no effect is assumed to be present below the NOAEL and where an expression of the critical effect is present above the NOAEL (see section 5.1 and Table 2). Given the typical animal studies used in toxicology, the effect size that can be detected by a statistical test may be larger than 10% (additional

risk). Therefore, the NOAEL may be expected to be a dose at which the effect is in reality somewhere between 0% and 10% or more. In contrast, the BMD is a dose for which the size of the effect has been predefined, and thus it is under the control of the risk assessor. Furthermore, while the dichotomy of the NOAEL approach does not provide quantitative information about risk above the ADI, such information might be derived from fitted dose–response models, where such dichotomy does not exist.

In general, DRM estimates are based on data from the entire dose–response curve for the critical effect. The standard NOAEL approach can be regarded as a special, simplified case of dose–response analysis, as it identifies a single dose that is assumed to be without an appreciable adverse effect. The dose–response model thus reflects the characteristics of the dose–response curve, particularly in providing estimates of slope. In the case of a regression framework, it provides standard error and confidence intervals for the model parameters.

5.2.2 Estimation procedure

NOAELs are restricted by the set of doses used in the specific studies. An important consequence is that the NOAEL may be either below or above the threshold it aims to approximate, assuming one exists. When the true threshold is higher than the NOAEL, the distance between the two can be expected to be limited (related to the dose spacing used). However, when the true threshold is lower than the NOAEL, the distance between the two is unlimited: the true threshold could be anywhere between zero and the NOAEL.

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

- *Group size.* The power to detect a NOAEL at some dose level is directly dependent on the sample sizes chosen at those dose levels (Gaylor, 1989). The larger the group size, the smaller the potential true effect size at the NOAEL.

- *Dose location*. Since the NOAEL is an applied dose that did not show significant effects, while the next higher dose applied did show significant effects, the NOAEL can only be one of the

doses actually applied in the study. A particularly disturbing disadvantage of the NOAEL approach is that in some cases no NOAEL can be assessed because the lowest applied dose showed effects.

- *Experimental variation*. Larger experimental variation between subjects will result in lower statistical power and, hence, higher NOAELs. In quantal data, this phenomenon is somewhat hidden, but in continuous data, it is directly visible: it is reflected by the scatter in the data per dose group. This experimental variation comprises various things: biological (e.g. genetic) variation between subjects, variation in experimental conditions (e.g. time of feeding, location in experimental room, time of section or interim measurements), and measurement errors.

DRM-derived estimates are based on interpolation, and these estimates are not restricted to the actually applied doses. DRM can also be used on a study where no NOAEL (only a LOAEL) can be defined, so in this situation another study may be unnecessary. A comparison of different models can be useful. When multiple models are fit to the same data and produce widely varying BMD estimates, caution should be used in interpreting the results, as this could indicate insufficient data for modelling (see chapter 6).

It should be noted that in comparison with the NOAEL approach, implementation of the full DRM approach may lead to differences between the NOAEL and the BMD in individual data sets. However, on average, BMDs that represent the lower confidence interval on the dosage giving a BMR of 5% or 10% tend to be quite similar to the NOAELs (Allen et al., 1994). Therefore, in data sets where DRM cannot be applied, the NOAEL may serve as a reasonable surrogate of the BMD.

5.2.3 Uncertainty

A modelling approach facilitates both sensitivity and uncertainty analyses. Uncertainty (see section 6.5) can be expressed numerically when the doses and responses are linked by a model. Such numerical analyses can also be subject to sensitivity analyses, to test the contribution of different aspects of the database or of

model characteristics to the overall uncertainty. The uncertainty in the risk estimates that arises from aspects of study design, such as dose spacing, sample size, and biological variability, can be assessed in a dose–response model. While uncertainty factors are amenable to uncertainty analysis (Slob & Pieters, 1998), the threshold procedure of the NOAEL is not readily amenable to quantitative estimation of uncertainty or to a sensitivity analysis.

A disadvantage of using the threshold procedure of the NOAEL for the estimation of a point of departure ("starting point") for formulating advice to risk managers is that it is not possible to quantify the degree of variability and uncertainty that may be present. The NOAEL is assumed to be a dose without biologically significant effects. This assumption is more likely to be valid in toxicological studies with larger sample sizes.

5.2.4 Study design

A design optimal for the NOAEL approach could limit the use of DRM, and vice versa. While the NOAEL approach requires sufficient sample sizes within dose groups (to warrant statistical power), the DRM approach requires a sufficient number of dose groups (to warrant a description of the whole dose–response). Given the restrictions on the total number of animals used in a single study, these two requirements may not be compatible.

An important point to bear in mind is that DRM can be used on studies carried out in the past and based on the traditional designs (with three dose groups and a control). Some have argued that optimal designs for dose–response models may have the advantage for animal welfare that fewer animals could be used (Slob et al., 2005).

While DRM provides uncertain estimates when the number of dose groups is too small, the determination of both the BMD/BMDL and the NOAEL may prove inadequate at different points when the number of animals per dose group is too small. For example, when the critical effect is seen in a larger experimental animal, such as the dog, with few animals per dose group, the NOAEL may be high owing to the insensitivity of the test. The BMD/BMDL approach, however, can be used to evaluate sparse dose–response data and quantify the inherent uncertainty. However,

even here, where an apparent dose–response relationship in the data remains, the BMD/BMDL may also provide very uncertain estimates. Therefore, a typical four-dose study with a few animals per dose may in practice be unreliable whatever method, NOAEL or BMD, is applied. However, the advantage of DRM is that this uncertainty is made visible, whereas in the NOAEL approach it remains hidden.

DRM reduces the need for more experiments when a small degree of extrapolation is needed (e.g. when the doses used are near the human exposure level). In contrast, the NOAEL approach may require further experiments where no clear NOAEL (or LOAEL) can be identified. This can be illustrated by the study of Allen et al. (1996) on developmental risk assessment in rats exposed to boric acid in their diet. This study failed to establish a NOAEL; however, the BMD approach could have been applied, thereby avoiding the need for repeat studies (see also section 5.2.2 above). Distributing the total number of animals over more dose groups does not result in poorer performance, despite the smaller number of animals per dose group, as shown by Slob et al. (2005). The above example of Allen et al. (1996) suggests that the BMD approach provides a reasonable basis for appropriately comparing and combining studies, as opposed to ad hoc combinations of study results.

A major advantage of DRM is the ability to estimate risks within the observable range of effects. In animal studies, it is possible to estimate risk over the full range of doses used. Estimation of risks outside the observable range will be more and more unreliable when risks get smaller and smaller (Murrell et al., 1998). Some studies (e.g. Sand et al., 2002) have investigated the effect of model dependence at different response levels.

Some experts have argued that extrapolation to risk levels outside the observable range might be warranted when there are indications that the same toxicological mechanism is active in both the extrapolation region and the experimental region of the model fit. However, mathematical models that adequately describe the full complexity of the mechanisms involved are very rare; even then, it needs to be additionally assumed that the parameters estimated from the data (i.e. the observable range) are adequate for the low-dose

range and have sufficient precision to make the prediction (see also section 6.5.3).

5.2.5 Biological information

The NOAEL approach incorporates biological information through the application of "expert" but subjective judgement. Full DRM has the potential for a more "science-rich" analysis through the more formal quantitative inclusion of factors/covariates into the models, in the case of both human epidemiological and animal data.

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

5.2.6 Comparison of experimental results

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

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

Rules for combining studies, however, need to be developed. Descriptions of the dose–response on the same end-points in different studies may be integrated to provide a cohesive picture of the chemical's toxicity. The values obtained using DRM may result in estimates for each end-point on the basis of biological and functional relevance.

5.2.7 Risk management perspectives

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

- giving the decision-maker a "more-than-one-point" appreciation of the data;
- providing information about what happens above the safety level (magnitude and types of health impacts);
- quantifying benefits in risk reduction from different regulatory actions;
- promoting consistency in decisions, if appropriate adjustments are made for differences in effect, effect level, species, and study design; and
- allowing for an iterative interaction between the risk assessor and risk manager on a continuous and ongoing basis.

5.3 Implementation issues

In the case of the BMD, there are a number of decisions to be made in applying the method and determining a BMDL: for example, which mathematical model to use; what degree of confidence to use in calculating confidence limits; what response level to predetermine as the BMR (e.g. BMR = 1%, 5%, or 10% incidence of an effect, or a 5% or 10% change in a continuous endpoint, such as body weight or red blood cell counts). It is often not clear what response level (BMR) can be considered as non-adverse. For example, should a 5% decrease in red blood cell counts be considered as adverse, or should a smaller (or larger) change be chosen? Should up to a 5% increased incidence in hepatocellular hypertrophy be considered as acceptable in an animal study, or is a maximum of 10% increase adequate? These and other choices need additional discussion among toxicologists and clinicians. Although an explicit statement on the BMR is an improvement compared with the generally unknown response level associated with a NOAEL, choices of a BMR need consensus building.

5.4 Summary

The characterization of dose–response relationships in animal and human studies has been a major component of hazard characterization. Over the years, a variety of methods have been developed to accommodate such relationships. DRM may be regarded as the most adequate approach for analysing dose–response data, provided that a suitable data set of animal or human dose–response data is available.

The standard NOAEL approach identifies a single dose that is assumed to be without appreciable effect, whereas the BMD is based on data from the entire dose–response curve, estimated for the critical effect. Although the effect at the NOAEL is assumed to be zero, it will be non-zero in many cases, although to what extent remains unknown. The size of the effect at the BMD is made explicit and, as far as possible, is based on toxicological knowledge. While the uncertainty in a NOAEL cannot be quantified, the uncertainty in a BMD can be quantified by a confidence interval. The use of DRM may call for different guidelines for optimal study designs, as the number of dose groups should be sufficiently large. Distributing the total number of animals over more dose groups may be done without loss of precision. DRM can more effectively compare different experiments, effects, and compounds. While risks above the ADI based on a NOAEL cannot be quantified, such may be possible for exposures exceeding the ADI based on DRM. For estimating risks below the observable range, extrapolation based on a single fitted model is unwarranted. Here, linear extrapolation may be considered as a conservative approach. Currently, more advanced methods are being developed (e.g. Bayesian approaches) for low-dose extrapolation based on DRM.

6. PRINCIPLES OF DOSE–RESPONSE MODELLING

6.1 Data

6.1.1 Selection of data

When considering which data to use from a set of available toxicity studies on a particular compound, it may not be effective to do a dose–response analysis for each observed end-point in each study. As a first step, one may omit studies that have obviously larger NOAELs compared with the other studies. In this way, one may, for example, select for a given type of toxic response (e.g. chronic, developmental) for the most sensitive species. For a given study, many end-points may have been measured. End-points not showing a clear dose–response on visual inspection can be omitted. Then, based on the toxicological impact together with the apparent magnitude of the response, a selection of end-points can be made as candidates for modelling. It would be very helpful if submitted studies included an annex with plots (in addition to tables) of observed data points for each end-point, possibly with fitted curves to the plots, to enhance the process of selecting end-points. At a minimum, these should be included for end-points showing evident effects.

After selecting the potentially relevant end-points, one must decide whether each dose–response data set is actually amenable for a dose–response analysis. Generally, it is desirable to have at least three or four different doses (including controls). In addition, the associated effect levels need to be different from each other; it is preferable to have at least three different response levels.

6.1.2 Data types

There are various types of response data, and these can be categorized in various ways. The main distinction relevant for effects is that between quantal and continuous data. Quantal data relate to an effect that is observed or not in each individual subject (laboratory animal or human). Hence, for each dose, the number of subjects responding out of the number of subjects available is

reported. In continuous data, a quantitative measurement is associated with each individual subject. As an intermediate type of data, ordinal data reflect (ordered) severity categories—that is, they are qualitative data but with a rank order (e.g. histopathological severity data). When the categories are non-ordered, they are called categorical data, but these are rare for response data. Finally, count data form another class of data (i.e. discrete data), but in practice they can often be treated as continuous data (see also section 4.3).

Although the type of data is important for statistical reasons (see section 6.2.2 on distributions), the distinction between quantal and continuous data also has a crucial impact on interpretation of results and their ensuing use in risk assessment. In the case of quantal dose–response data, information on the change of incidence with dose is available at one particular degree of effect. For example, the incidence of cleft palate may increase as dose increases, but under the categories "no cleft palate" and "cleft palate", there is no information about the degree of the effect. In ordinal and continuous data, in contrast, information on both the degree of effect and the incidence is available as a function of dose. So, for example, cleft palate might be categorized into an ordinal variable with levels "no clefting", "mild clefting", "moderate clefting", and "severe clefting", or it might be quantified in a continuous variable as, for example, the fraction of closure. The relationship between the average response and dose gives information about how exposure changes the degree of effect. For instance, a plot of (average) red blood cell count may show the decrease in mean red blood cell count (i.e. the degree of the effect) as a function of dose. By also considering the individual data points, information on the incidence can be derived as well. For example, an estimate of the fraction of individuals with red blood cell counts less than some critical value can be derived.

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

6.2 Models and distributions

6.2.1 Dose–response models

6.2.1.1 Continuous dose–response models

The models listed in Table 4 are some of the forms that may be used to describe the relationship between dose and the magnitude of a response on a continuous scale in an individual. When combined with a statistical distribution (e.g. normal or lognormal), these equations can also be used to describe the relationship between dose and a continuous response in a population, where the continuous model corresponds to the central estimate.

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

1. $y = a + f_x(D)$
2. $y = a \times f_x(D)$
3. $y = f_x(a + D)$

where D is dose, a is the background term, and f_x may be any dose–response function. For some assessments, there may be mechanistic information that makes one form preferable to another. For example, the first form is preferable for modelling an influence that produces the effect independently, the second corresponds to the idea of normalizing the response as a fraction of the background response, and the third reflects a contribution from another agent acting by the same mechanism.

Table 4. 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).	$= \mathrm{RMax}\frac{[S]}{K_M + [S]}$	RMax is the maximum rate of the reaction, [S] is the substrate concentration, and K_M is the Michaelis-Menten constant, which 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.	$= \mathrm{RMax}\ \frac{D^n}{K_D{}^n + D^n}$	RMax is the maximum response, D is the dose, K_D is the reaction constant for the drug–receptor interaction, and n is the number of (hypothetical) binding sites.
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.	$= \mathrm{RMax}\left(1 - e^{-rD}\right)$	RMax is the maximum response, D is the dose, and r is the exponential rate constant.
Power	Simple exponential model.	$= \beta D^{\alpha}$	D is the dose, α is the shape parameter, and β 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.	$= mD$	D is the dose and m is the slope.

6.2.1.2 Quantal dose–response models

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

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

1. $y = a + (1 - a)\, f(x)$
2. $y = f(x + a)$

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

6.2.1.3 Thresholds

The term “threshold” can be used in three different senses. First, it is used in a scientific sense to indicate a level of exposure at which no effect occurs (e.g. there is a physical stimulus, but there is no response). Second, a threshold may be thought of as a level at which there may or may not be an effect, but it is too small to be observed (e.g. a NOAEL). In this case, it is the perceptual limitation of an observer or analyst, rather than the actual subject of the experiment, that is being described. As a third meaning, a “practical

Table 5. Quantal dose–response models

Name(s)	Theoretical basis	Equation for frequency (F)	Parameter explanations
Step function	No variability.	If $D < T$, $F = 0$ If $D \geq T$, $F = 1$	D is the dose and T is the threshold parameter.
One-hit (single-hit)	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).	$= 1 - e^{-(\alpha+\beta D)}$	D is the dose, e is Euler's constant, α is a location parameter, and β is the slope parameter.
Gamma multi-hit	An expansion of the one-hit model, which is based on the notion that multiple hits or events are required to produce a particular effect.	$= \Gamma(gamma*D, k)$	Γ() is the incomplete gamma CDF, D is the dose, gamma is a rate parameter, and k is the number of hits required to produce the effect.
Probit normal	A descriptive model based on a normal or Gaussian distribution.	$= \Phi(\alpha + D*\beta)$	Φ() is the normal CDF, D is the dose, α is a location parameter, and β is the slope parameter.
Logistic	The statistical logistic model is also a descriptive tool with no theoretical basis.	$= \frac{1}{1 + e^{-\alpha - D \times \beta}}$	D is the dose, α is a location parameter, and β is the slope parameter.
Weibull	A flexible descriptive model originally developed to describe survival data in demography.	$= e^{-(\alpha + (\beta \times D)^{\gamma})}$	D is the dose, α is the background parameter, β is the slope parameter, and γ is an exponent.

CDF, cumulative distribution function

threshold" is a response where the consequences are determined to be trivial and not worth further consideration.

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

- Below threshold, the effective dose is zero.
- Above threshold, the effective dose is the dose minus threshold.

Threshold terms generally are difficult to estimate accurately and have large confidence limits.

6.2.1.4 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 end-point. As noted above, the dose–response of continuous end-points may be directly interpreted as a dose-related change in degree of effect—for example, a per cent decrease in haematocrit (Woutersen et al., 2001) or a per cent change in body weight (see Figure 5, which represents the dose–response relationship between body weight and exposure to the mycotoxin deoxynivalenol) (Pieters et al., 2004). Here, a certain degree of effect (5% reduction in body weight) is chosen for deriving the BMD. The BMDL is then defined as the dose associated with a particular (e.g. 5%) change in degree of effect for that end-point.

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 needed only to a small extent when continuous end-points are considered.

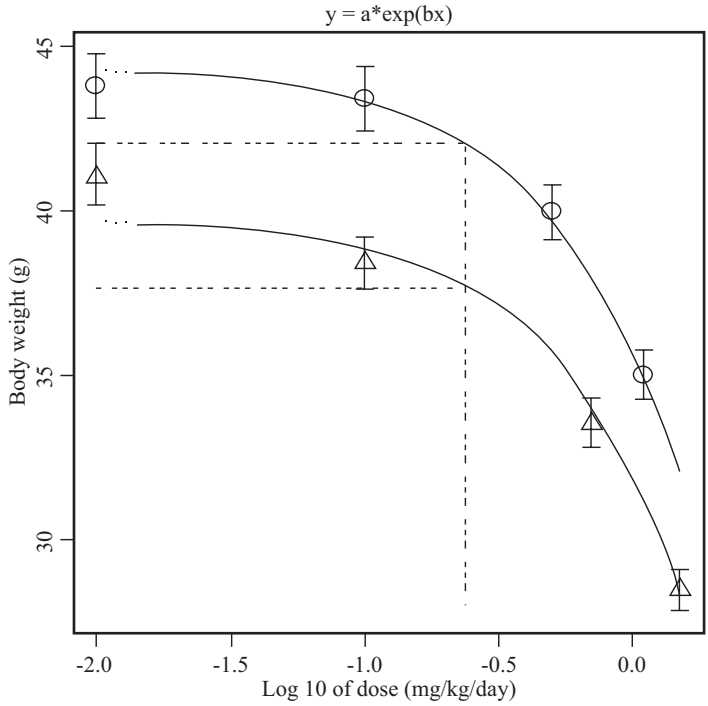


Fig. 5. Dose–response model fitted to male (circles) and female (triangles) body weights plotted against log dose (exposure to the mycotoxin deoxynivalenol). The plotted marks represent the (geometric) means of about 40 mice, with 90% confidence intervals. The BMD associated with a BMR of 5% is estimated at 0.24 mg/kg body weight (log equivalent = −0.62), with a lower confidence bound of 0.22 mg/kg body weight (log equivalent = −0.66). The latter value can be considered as a BMDL for this end-point (adapted from Pieters et al., 2004).

In the case of a histopathological end-point resulting in ordinal data, a dose–response function may be fit 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—that is, in an analysis of multiple studies (Hertzberg & Miller, 1985; Hertzberg, 1991; Hertzberg & Wymer, 1991). In this application of categorical regression, severity categories are defined covering disparate end-points. Most of these applications focus on estimating

the likelihood that a given category of severity may occur at a given dose level.

6.2.1.5 Modelling with covariates

In some circumstances, it is desirable to include variables in addition to an exposure variable in dose–response models. For example, in epidemiological studies, it is common to model disease risk in terms of not only exposure, but also age, sex, socioeconomic status, smoking status, and other measurements that may be relevant to the disease state. These other factors may be correlated with exposure status because of the way in which the sample was taken. Then, unless the proper covariates are included in a model for the relationship between exposure and the health end-point, the effect of exposure will be incorrectly estimated. In bioassay studies, in which animals are randomized to treatment groups, this sort of confounding cannot, in principle, occur, but it may be useful to include a covariate such as sex to account for some of the variability in a related measure (see Figure 6).

6.2.1.6 Biologically based dose–response models

While biological considerations may motivate the choice of one or several empirical models, the level of biological detail in such models is minimal. Thus, their credibility for interpolating and extrapolating a data set derives mainly from their fit to the data, as evaluated statistically. Another class of model, the biologically based dose–response model, is much more complicated and is explicitly designed to model the biological details that lead from initial exposure to a toxicant to the ultimate pathological outcome. Typically, such a model includes a physiologically based toxicokinetic model to describe the distribution and metabolism of the parent compound and toxic metabolites and other mechanistic, or toxicodynamic, models that link target tissue concentration to the ultimate response. The toxicodynamic part of the models may be relatively simple (e.g. when the outcome is inhibition of acetylcholinesterase in the model for chlorpyrifos; Timchalk et al., 2002) or as complicated as a fully elaborated stochastic model for carcinogenesis (Sherman & Portier, 1998). Such a model is really a quantitative expression of a set of biological hypotheses and, when rigorously tested against critical experiments, becomes a credible tool for extrapolating from experimental results into exposure

realms that are difficult or expensive to reproduce in controlled experiments. Such models are quite expensive to construct both in resources and in time and thus would be expected to be developed fully only for exposures and toxicities of the highest concern.

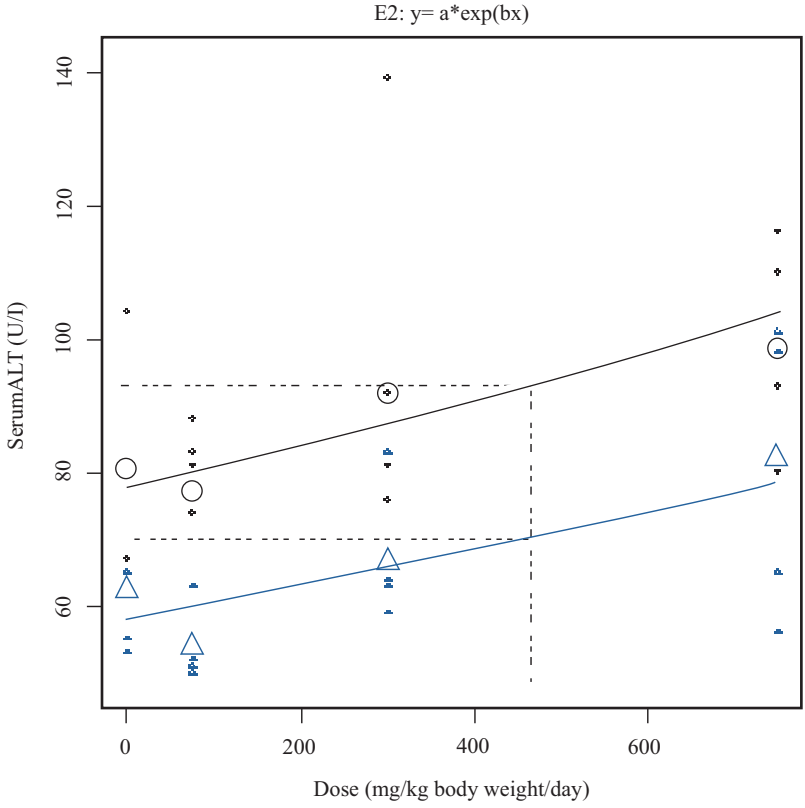


Fig. 6. Dose–response model fit to serum alanine aminotransferase (ALT) levels observed in males (circles) and females (triangles), where sex is treated as a covariate. In this case, the parameter a (background response level) differs between sexes, whereas parameter b and the residual variance (var) for the log(data) do not differ between sexes.

6.2.2 Statistical distributions

6.2.2.1 Continuous distributions

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

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

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

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

6.2.2.2 Discrete distributions

Discrete distributions describe responses on a finite or infinite scale, preferably count data; a special case is a response with a dichotomous quantal outcome of 0 or 1.

A Bernoulli distribution has an outcome of 1 or 0, corresponding to the occurrence or absence of an event that 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 subjects in which the specified event occurs 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 events per trial is small. Consequently, the Poisson distribution 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 incidence 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).

6.3 Model fitting and estimation of parameters

The general principles of parameter estimation and model fitting have been discussed in chapter 4. 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 non-linear modelling in general has been restricted to using conventional methodologies, whereas 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 & Jarabek, 1995; Gelman et al., 2004). While such software requires substantial statistical understanding for successful use of Bayesian methods and is thus beyond the scope of this document, even conventional methods require an understanding of some basic principles before outcomes from applying the software can be properly interpreted. Some general remarks are given below.

6.3.1 Criterion function

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

In many applications, the logarithm of the likelihood function is used as the criterion. The likelihood directly derives from the distribution assumed for the scatter in the data. For quantal data, the binomial likelihood is typically used. For continuous data, the normal likelihood is often used, be it for the observed responses themselves or for the log-transformed responses. Note that maximizing the likelihood function for data that are assumed to be

normally distributed is in fact equivalent to minimizing the sum of squares.

6.3.2 Search algorithms

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

An iterative search algorithm tries to find "better" parameter values in a process by evaluating whether the fit can be improved by changing the parameter values through a trial and error process. More advanced algorithms operate by evaluating the slope of the likelihood at which the fit is improved for one or more parameter value changes (basically using the slope to "climb the likelihood function" as quickly as possible to find the top value). The algorithm can start searching only when the parameters have values to start with. Although the software often gives a reasonable first guess for the starting values, the user may have to change these. It is not unusual (in particular when the information in the data is hardly sufficient to estimate the intended parameters) that the end result depends on the starting values chosen, and the user should be aware of that.

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

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

2. 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 dose–response model with five unknown parameters cannot be estimated with a study with four dose groups. As another example, the variation between the observations within dose groups may be large compared with the overall change in the dose–response. 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.

6.4 Model comparison

The fundamental criterion for judging a model is that the selected model should describe the data, especially in regions of the dose–response 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 a p-value of 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 would be extremely helpful for plots that include data points to also include a measure of dispersion of those data points. In certain cases, the typical models used in DRM cannot fit the observed data, such as 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 transformation of dose) or the model (e.g. adjustments for unrelated deaths) may be helpful.

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 incorporating 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 $2 \times (LL_{full} - LL_{restricted})$, where LL is log-likelihood, and compare this with a critical value from the chi-squared distribution with $P_{full} - P_{restricted}$ degrees of freedom (where P_x is the number of parameters estimated in model x).

2. The models are from the same family, but do not form a nested series. Some statistics, notably Akaike's information criterion (AIC is $-2LL + 2P$, where LL is the log-likelihood at the maximum likelihood estimates for the parameters, and P is the number of model degrees of freedom) can be used to compare models (Akaike, 1973; Burnham & Anderson, 2002). In this case, the model with the smallest AIC value is selected, although models with similar AIC values (differing by no more than about 4) are probably equivalent (Burnham & Anderson, 2002).

3. The models are not from the same family, but are fit using the same assumptions about the underlying probability distributions (e.g. all using a lognormal likelihood or all using a normal likelihood). In this case, Burnham & Anderson (2002) argue that AIC can still be used to identify the best model, but this appears to be a controversial point. Sand et al. (2002) have shown that it may be difficult to discriminate between the commonly used quantal dose–response models based on the AIC, which may be due to the fact that these models are quite

similar in their structure and include a similar number of parameters. In general, this case is still the subject of statistical research. At present, it will probably be adequate to use AIC to select a model as in the previous case, recognizing that this guidance may change.

4. Models do not use the same probability distribution. In this case, little formal statistical guidance is available. The plausibility of assumptions about the distribution of data needs to be examined by looking at the distribution of individual data. However, continuous data are often aggregated and reported as means and standard deviations, which eliminates the possibility of examining distributional assumptions. In these situations, the best that can be done is to rely on past experience with the end-points being modelled and select a reasonable probability distribution.

6.5 Representing uncertainty

Any parameters or predictions estimated from a given model are only point estimates and, to a larger or smaller extent, uncertain. This uncertainty arises from at least three sources:

1. *Sampling error*—the sampling error arising from inferences about a larger population from a single experiment;
2. *Study error*—the reality that dose–response estimates often differ among experiments with different experimental design, protocol, or uncontrolled circumstances; and
3. *Model error*—the fact that the "true" model is not known, which results in additional uncertainty when interpolating between doses, but even more so when extrapolating outside the dose range containing observations.

These three sources of uncertainty are briefly discussed below.

6.5.1 Sampling error

Uncertainty arising from sampling error with a single experiment is perhaps the easiest to evaluate and report. It may typically be quantified by a standard error or, preferably, by a

confidence interval. Confidence intervals may be calculated in several ways:

- plus or minus twice the parameter's standard error (provided by most dose–response software), which is estimated by the second derivative of the likelihood function (Hessian or information matrix);
- based on the profile of the log-likelihood function, using the chi-square approximation of the log-likelihood;
- bootstrap methods (see, for example, Efron, 1987; Efron & Tibshirani, 1993); and
- Bayesian methods, in particular if one has some preliminary knowledge of the plausible range of the parameter.

Various studies have compared the first three methods and concluded that the first may result in inaccurate intervals, whereas the second and third methods give similar results (see, for example, Moerbeek et al., 2004).

6.5.2 Study error

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

6.5.3 Model error

The third area of uncertainty, model uncertainty, is reflected by the question: to what extent do the data, possibly along with other knowledge about dose–response shape, constrain the set of possible dose–response shapes? A statistical model completely hinges on the

dose–response data, and the quality of the data is in fact the crucial aspect. In the fitting process, a model tries to hit the response at the observed doses. However, when a model is used to make inferences, interpolation between observed doses and extrapolation beyond the non-control doses are possible approaches. Thus, the model must also predict the response in the non-observed dose range. In other words, there are two aspects in evaluating the fitted model: one should assess not only if the model succeeded in describing the observed responses, but also if the model can be trusted to describe the non-observed responses where it is desirable to make inferences. The former aspect focuses on the quality of the model, the latter on the quality of the data. The following discussion elaborates on how to deal with the second of these two aspects (the first was addressed in section 6.4, Model comparison).

There are two ways to evaluate whether the data provide sufficient information to constrain the model and allow inference in some defined range outside of the range of the data. The fitted dose–response model should be visually inspected, to check if the data provide sufficient information to confine the model. Here, the question should be asked: if a curve is drawn through the data points by hand, could that be done in disparate ways? For instance, in the top panel of Figure 7, three curves have been drawn through the data points, each of which might be close to the true dose–response curve in the range between 2 and 5. In the bottom panel, however, it is very difficult to imagine that the true dose–response relationship would be very different from the (single) curve drawn here in the same range.

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

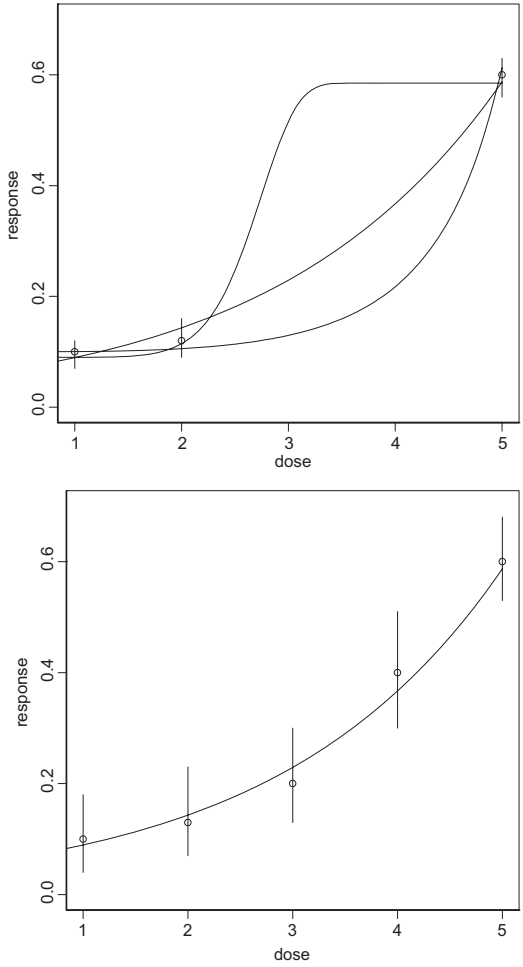


Fig. 7. Two data sets illustrating the idea of model uncertainty. In the top panel, the data (either quantal or continuous) do not contain sufficient information to confine the dose–response relationship in the range between 2 and 5: one may imagine various disparate curves that are all in agreement with the data, and hence they all might represent the true dose–response relationship. In the bottom panel, the data points prohibit the possibility of drawing disparate curves between 2 and 5.

In current practice, there is a tendency to focus only on goodness of fit, and passing a formal goodness-of-fit test is often

regarded as sufficient evidence that the model is acceptable. This is unfortunate, since a goodness-of-fit test tends to be more easily passed for data with few dose groups or when few dose-related responses are noted and therefore non-observed responses are important or dominate. In addition, a goodness-of-fit test assumes that the experiment was carried out perfectly (i.e. perfectly random with respect to all potentially relevant experimental factors and actions). Clearly, this assumption is not realistic.

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

Dose–response models that are based on the mechanism of action of a 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 suggested 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, as shown in the top half of Figure 9, and so give similar predictions in the range of the data, but whose predictions diverge at the low end of the dose range, as depicted in the lower half of Figure 9. One way to collect and represent model uncertainty in a risk assessment is through the

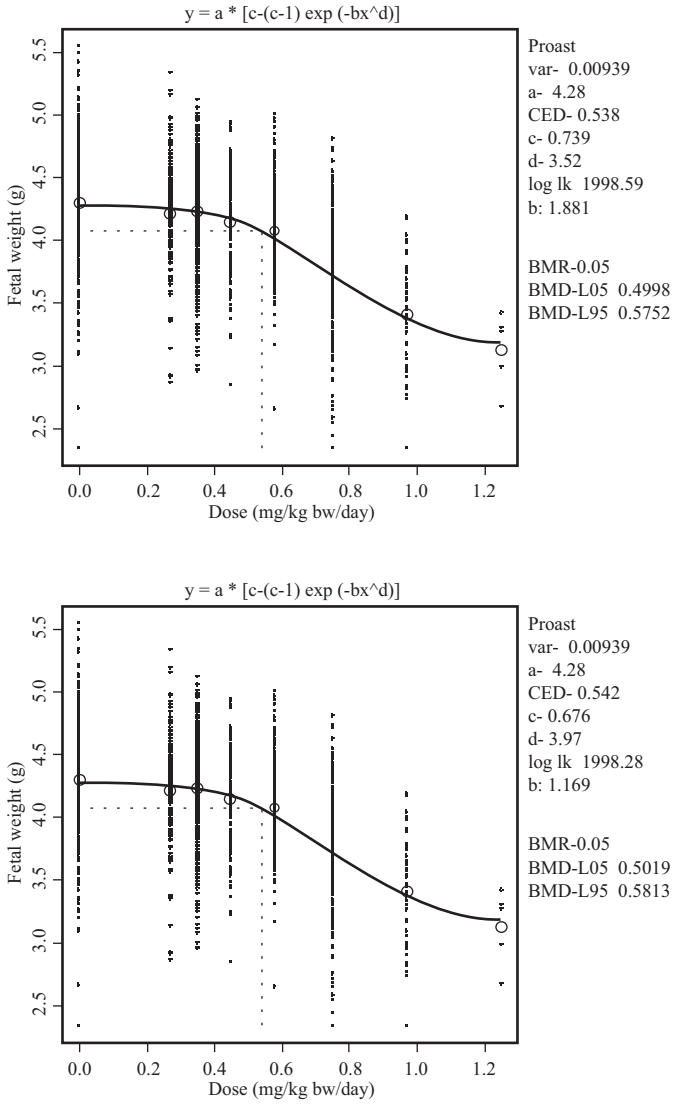


Fig. 8. Two different models (both with four parameters) fitted to the same data set resulting in similar dose–response relationships and similar BMD(L)s. Small circles indicate individual observations, large circles (geometric) group means.

use of probability trees (Rescher, 1969; 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 top panel of Figure 8 were equally well supported, then each model would have a probability of 0.33. If one model had a weight that was 6 times greater than the others, then it would have a probability of 0.75, whereas 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 judgement. 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 at 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 semiformal approach may be used in which the same criteria discussed in the section for selecting models (section 6.2.1) may also be used to weight and assign probabilities to each alternative model considered (e.g. Carrington & 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 models from each bootstrap are then retained in a probability tree to represent both parameter and model uncertainty.

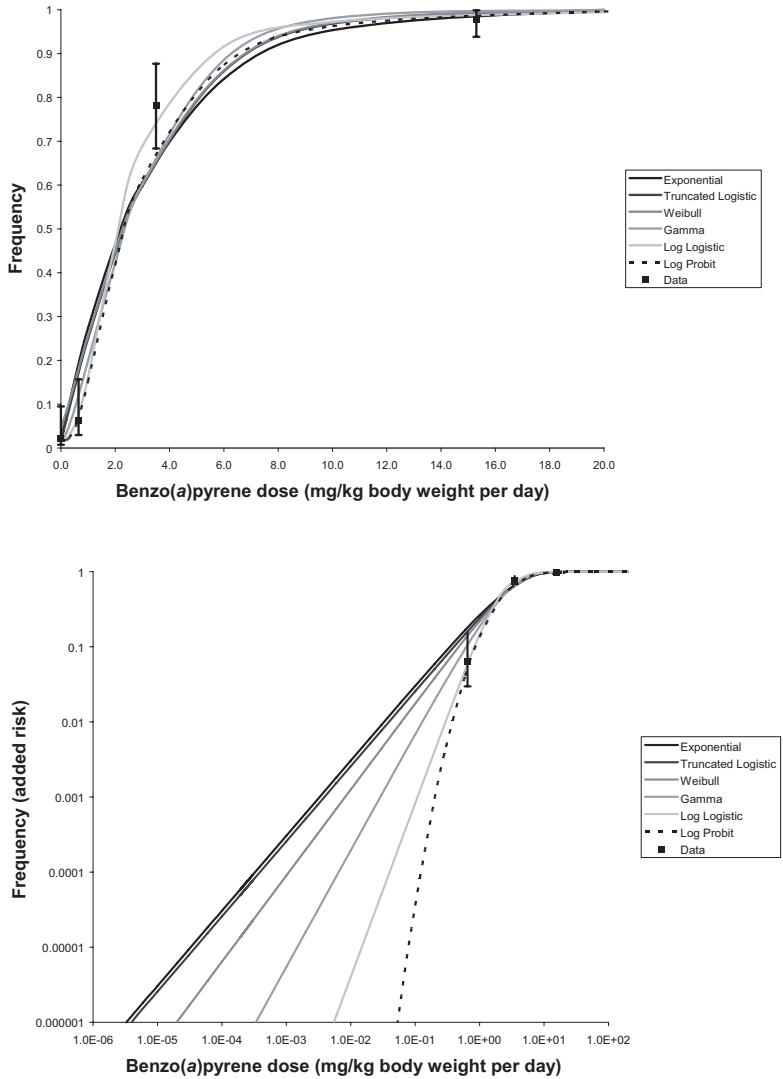


Fig. 9. Model uncertainty in low-dose extrapolation. Different models may all fit the data reasonably well (top), but yield highly divergent response estimates at low doses (bottom). The data and models are taken from Fitzgerald et al. (2004).

Alternatively, some people have addressed this uncertainty by choosing a subset of the models that appear to fit the data well. From these models, those with adequate fits are summarized with a range and associated variance. When choosing a final value for the BMD, these values can be aggregated by taking a mean or geometric mean to provide a central point estimate (National Health and Medical Research Council, 1999) or a value simply chosen through expert judgement (WHO, 2006).

6.6 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 ADIs. The main advantages of the use of the BMD over NOAELs and LOAELs stems from the more complete use of 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 BMR is the response for which the BMD is to be calculated. There are both technical and policy aspects associated with selecting the BMR. The technical aspects have to do with just how the BMR is expressed; different types of end-points, 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 it does not directly address the choice of its particular value.

The way in which the BMR is expressed depends upon the kind of response variable being modelled. For end-points 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 (AR):

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

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

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

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 (RR):

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

BMRs for continuous end-points 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 acetylcholinesterase activity is inhibited by 10% relative to control levels (this is often termed the critical effect size). 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 end-points that demonstrate a sigmoidal response, as does enzyme induction, it has been suggested (Murrell et al., 1998; see Gaylor & Aylward, 2004, for a contrary argument) that a formulation similar to extra risk be used: for these end-points, 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 (e.g. the difference between background and maximum possible induction) of the response. The Gaylor & Aylward (2004) approach considers a certain setting within the definition of the response (i.e. a 1% change) and compares the uncertainties in the resulting BMD with the uncertainties in BMDs estimated using the specific setting in the "hybrid" approach. Thus, their conclusion may not hold in general terms (e.g. considering a 5% or 10% change in response relative to the total dynamic range).

Indirect or "hybrid" approaches have been advocated by Crump (2002) and Gaylor and his co-authors (Gaylor & 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 that 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 interindividual 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. Sand et al. (2003) examined how the hybrid approach depends on the estimate of variance. Gaylor & Slikker (2004) discussed how different sources of variability may be separated.

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 6.2.1.4). In developmental bioassays, characteristics of the dam or the litter as a whole (e.g. 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 sex. 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, a detailed analysis of the dependency should be undertaken (e.g. modelling the BMD as a function of that covariate). If the variable makes sense for extrapolation to the human situation, it might be informative to calculate the BMDs for several values of the covariate, to evaluate the sensitivity of the BMD to the range of covariate values for humans.

6.7 Summary

Data sets for DRM generally need to be selected to reflect the more sensitive end-points 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 dose–response 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; in addition, their ability to describe the data with respect to each other may be compared using measures such as the AIC.

Uncertainty about the inferences that result from such models 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 dose–response relationship. Dose–response analysis needs to address all three sources of uncertainty whenever possible.

One particularly important application of DRM is the calculation of BMDs, doses at which it is inferred that a particular level of response would occur. When data are available, BMDs are a better alternative than 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.

7. COMMUNICATING THE RESULTS OF DOSE–RESPONSE MODELLING

7.1 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" (IPCS, 2004). Risk communication has evolved with the rest of the risk analysis paradigm to embrace the "interactive" nature of the processes. The transition from monologue to reflexive dialogue 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 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 marginalization 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 level, 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 health-based guidance value 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 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 limitations and weaknesses of the technical knowledge in addition to its strengths. There should also 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 processes of risk assessment and risk management and thereby facilitate communication. As a result, participation in the decision process will be broadened.

7.2 Incorporation of the outputs of dose–response modelling into risk assessment

The output of dose–response analysis can be used in various ways, depending 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:

1. establishment of a health-based guidance value, such as an ADI or TDI, which is a daily intake over a lifetime that is considered to be without appreciable health risk (this would be analogous to current procedures based on a NOAEL or LOAEL);
2. estimation of the MOE as the ratio between the dose–response output and the estimate of human exposure; and
3. 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 dose–response model was the external dose expressed in milligrams per kilogram body weight. The use of internal or target organ dose estimated by a physiologically based toxicokinetic 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.

7.3 Derivation of health-based guidance values

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 CSAFs (IPCS, 2005), 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 depends on the sample size, and more than 100 animals may be needed to achieve confidence intervals in the range of ±5%.

Many studies have shown that the BMDL for a 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 with BMDs 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 dose–response 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. This possibly can be explained by the smaller sample sizes in these experiments, not by the difference in end-points.

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.

7.4 Estimation of the margin of exposure

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, 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, 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 margin or safety and is therefore termed an MOE. The MOE is calculated as the ratio between two experimental estimates, the BMDL and the predicted or estimated human intake/exposure. Calculation of an MOE does not require extrapolation of the data beyond the range of observations (IPCS, 1999; 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 one in a million lifetime risk is equivalent to an MOE of 50 000.

7.5 Quantitative estimations of the magnitude of the risk at levels of human exposure

The results of a dose–response 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 dose–response 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 dose–response relationship. 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. However, if one assumes that the dose–response relationship in humans has a similar shape to that in the animal study, the ADI is set with default uncertainty factors that will obscure any quantitative estimates of the risk above the ADI (the risk at the ADI is assumed to be negligible). More accurate estimates of differences in sensitivity between humans and animals would be required for such calculations.

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 dose–response 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 dose–response relationship defined by the model that was 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 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.

7.6 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.

7.6.1 Tables

Precise communication of quantitative information requires numbers. More numbers will portray more information than fewer numbers, but will take longer to assimilate. Tables 6–10 give examples of the range of options, from high to low complexity, that may be considered, all taken from the same simulation results for exposure. It is recommended that, in case of effects of concern or a single effect found in several studies, all quantitative results be summarized in a table. The risk assessor should sort out the most relevant results and present the data to the risk manager in a clear and understandable way.

7.6.2 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 10). The first is to plot density against value, which emphasizes the values that are the most common or likely. The second is to plot cumulative percentiles against 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 which dose descriptor should be used for lifetime risk.

Table 6. Population percentiles from a two-dimensional simulation

		Uncertainty												
		Average	SD	Minimum	P1	P5	P10	P25	Median	P75	P90	P95	P99	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
	P1	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
	P5	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
	P10	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
	P25	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
	P75	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
	P90	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
	P95	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
	P99	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

Pxx = xxth percentile; SD = standard deviation.

Table 7. Population percentiles with confidence intervals

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

Table 8. Population percentiles with standard deviations

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

Table 9. Selected population percentiles with confidence intervals

Percentile	Average (confidence interval)
Average	0.457 (0.366, 0.503)
Median	0.401 (0.267, 0.476)
90th percentile	0.808 (0.774, 0.848)
95th percentile	0.949 (0.930, 1.014)
99th percentile	1.247 (1.142, 1.403)

Table 10. Population mean with uncertainty estimate

	Average ± standard deviation
Average	0.457 ± 0.063

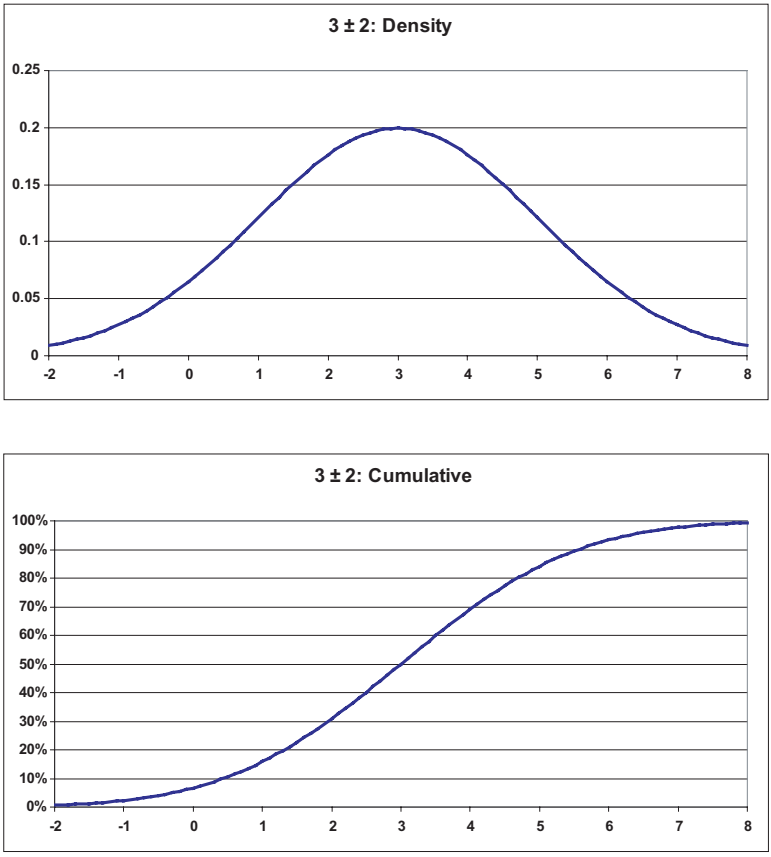
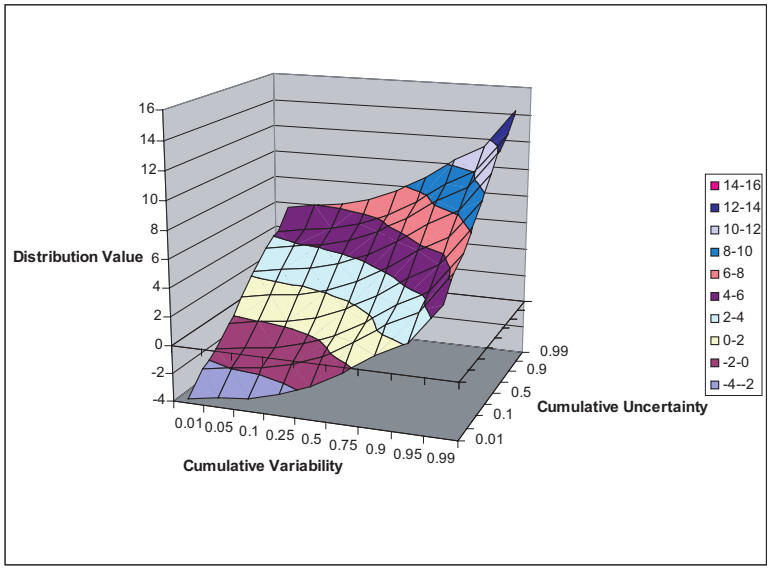


Fig. 10. Plotting frequency distributions.

Two-dimensional results are more difficult to display. Two strategies for adding an extra dimension are illustrated in Figure 11. 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.



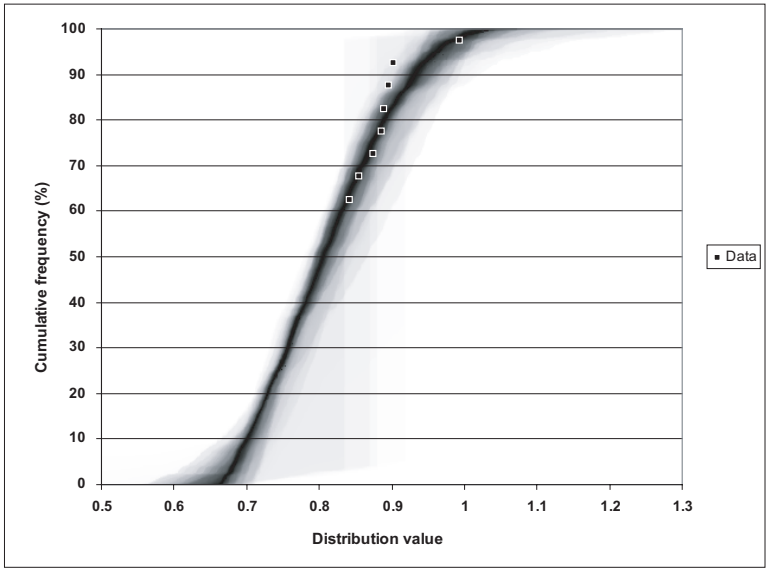


Fig. 11. Plotting results of three-dimensional simulations.

7.7 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 risk assessment procedures that have been accepted internationally as the indicator for determining acceptable levels of exposure. These rely on the identification of a NOAEL/no-observed-effect level (NOEL) for a critical end-point in the effect data and incorporation of uncertainty factors to allow for interspecies and interindividual 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 assessments 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 the (critical) effect;
- limitations in the interpretation;
- assumptions made in the analysis; and
- qualitative assessment of the potential effects of exceeding the health-based guidance value.

7.8 Synopsis of approach to modelling

DRM involves six basic steps: data selection, model selection, statistical linkage, parameter estimation, implementation, and evaluation (see chapter 4, Table 1). In undertaking a DRM exercise,

two factors that will impact the types of outputs and that may be of importance to the risk manager are briefly described below.

7.8.1 Data sets

Traditional safety assessments focus primarily on a single critical end-point, whereas DRM 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: namely, 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 data set is presented, will need guidance on the rationale for forwarding the additional data set information and for synthesizing this additional information. This guidance may include information about the consistency (or inconsistency) of the quantitative response across the end-points. Such information could be used by the risk manager to strengthen (or weaken) his or her confidence in the quantitative evaluation of the potential for health impacts.

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

7.8.2 Uncertainty

DRM should capture the relative uncertainties in the estimates of risk. This information will allow the generation of confidence limits on health-based guidance values. However, such confidence limits 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 levels. Likewise, dose–response functions have also been utilized to calculate percentiles of the population above target risk levels.

One of the advantages of DRM is that the confidence limit around the BMD can be calculated. From the conservative point of view, the lower limit of the dose is most important. However, this is not the same as to say that the confidence limit of the health-based guidance values can be calculated, as the uncertainty factors will obscure such estimates.

7.9 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 dose–response 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.

7.9.1 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 dose–response 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 (e.g. 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 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.

7.9.1.1 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 7.3). 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 that 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 among risk management options.

7.9.1.2 Margin of exposure

An MOE is determined by comparing the point of departure (the BMDL) with 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 owing to the nature of the effect, such as for substances that are genotoxic and carcinogenic.

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 dose–response 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 an 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 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 (National Health and Medical Research Council, 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 <5000, 5000–500 000, and >500 000 indicate high, medium, and low priority, respectively.

- The Committee on Carcinogenicity in the United Kingdom considered derivation of the minimal risk level for a genotoxic and carcinogen compound. One proposal was that an adequate MOE for carcinogenicity might be 10 000 (Gaylor et al., 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 (BMD_5) (Committee on Carcinogenicity, 2003). The MOE for average intakes for acrylamide in men in Norway has been estimated using the T_{25} value[1] and the LED_{10} (the lower bound on the effective dose for a 10% increase in risk) (approximately equivalent to $BMDL_{10}$) methods. These approaches result in MOE values of 1306 and 1225 for T_{25} and LED_{10}, respectively.

[1] The tumorigenic descriptor T_{25} is the chronic daily dose that will give 25% of the animals tumours above background at a specific tissue site. The T_{25} is determined by linear interpolation from the lowest dose giving a statistically significant increase in tumours (Dybing et al., 1997).

- The 64th (WHO, 2006) and 67th (WHO, 2007a) meetings of JECFA used MOE approaches for the evaluation of several substances that were genotoxic and carcinogenic. The 64th JECFA developed general considerations for the formulation of advice on compounds that are both genotoxic and carcinogenic. This meeting established MOEs for acrylamide, ethyl carbamate, polybrominated diphenyl ethers, and polycyclic aromatic hydrocarbons. The 67th JECFA established an MOE for 1,3-dichloro-2-propanol.

- A joint European Food Safety Authority/WHO conference on the risk assessment of substances that are both genotoxic and carcinogenic (Barlow et al., 2006) compared the approaches that are currently used. "This conference concluded that the MOE approach was a useful and pragmatic option.…"

- O'Brien et al. (2006) presented a critical appraisal of the approaches to the risk assessment of genotoxic carcinogens in food and concluded that "Overall, MOE is the most appropriate default approach because it combines information on potency and exposure, without the generation of numerical risk estimates of unknown reliability." They presented case-studies on the calculation of MOEs for acrylamide, aflatoxin B_1, benzo(*a*)pyrene, dimethylnitrosamine, ethyl carbamate, and 2-amino-1-methyl-6-phenylimidazo(4,5*b*)pyridine.

7.9.2 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, the prediction of risks at specified exposure levels and functions that allow the estimation of exposure levels resulting in specified risks. In addition, DRM exercises can provide uncertainty analyses. The availability 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 United States and Europe for non-threshold (genotoxic) carcinogens. In the area of food safety, the United States Food and Drug Administration 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. 1–10% 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 linearized multistage model was previously extensively used by the United States Environmental Protection Agency (USEPA, 1986). The LED_{10} method was later proposed by the USEPA (1996), and the T_{25} (Dybing et al., 1997; Sanner et al., 2001) method has been used in Europe (European Commission, 1999; SCCNFP, 2003). Lifetime cancer hazards may be estimated by linear extrapolation using LED_{10} and T_{25} as starting points. 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 scenario of human exposure to such substances, with some default policy-determined risk level.

7.9.2.1 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, the number of excess skin cancers from arsenic-contaminated water, and the 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 as "Up to X 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 may be; 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 interpretation by the risk manager include choice of models, choice of end-points, and limitations in initial data sets that were extrapolated.

One use of such information has been to evaluate the effect of different maximum limits for a chemical on risks. This type of consideration was included when JECFA evaluated aflatoxin B_1 and the impact of different maximum limits on risk (WHO, 1999, 2007b). Similar assessments have also been performed for lead and fumonisins B_1 and B_2 (Carrington et al., 1996; Humphreys et al., 2001). For example, the health impacts of current particulate standards (WHO, 2000, 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.

7.9.2.2 *Prediction of exposure levels producing specified 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 (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.

7.9.2.3 *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 of population effects with confidence intervals. The result from such analysis is a distribution of potential population risks. For example, in Figure 12, the outputs for three models and two data sets were used to generate a set of 3000 different model parameters (two data sets, three models, 500 bootstraps).

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 1–2 orders of magnitude closer to environmentally relevant exposures. Such approaches can be facilitated when DRM data are available.

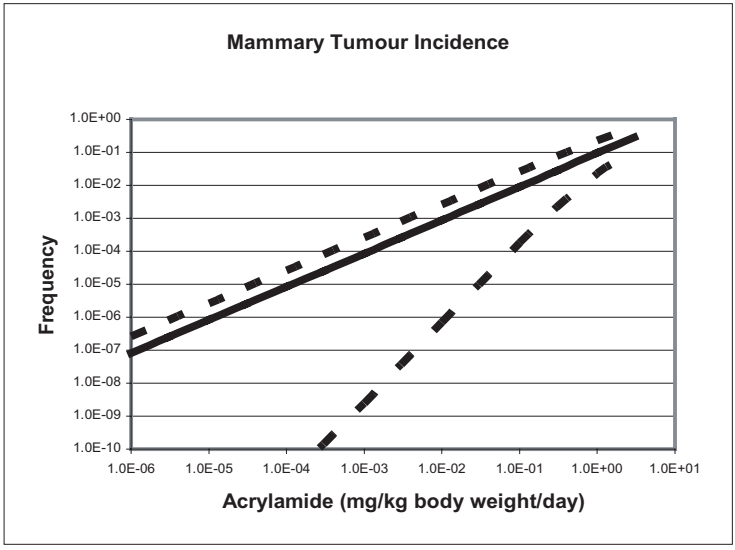


Fig. 12. Integrated uncertainty analysis for mammary tumours. The dark line is the central (median) estimate, and the dotted lines are the 5% and 95% confidence limits.

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 cost–benefit and risk–benefit analyses. The limitation is the question of whether the model outputs are accurate and representative of public health impacts.

7.10 Issues for risk managers

7.10.1 Risk assessment issues

7.10.1.1 Population versus individual effects

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

The output of the dose–response model should identify the degree of any subpopulation effects.

7.10.1.2 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 dose–response model 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. IPCS, 2000; 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.

7.10.2 Risk management issues

7.10.2.1 Risk management options

A risk assessment can be used to establish that a risk is of a sufficient magnitude that regulation or other type of intervention may be warranted. DRM can 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 49th JECFA analysed the application of two hypothetical standards for aflatoxin contamination in food in model populations (WHO, 1999).

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. aflatoxins), advice on consumption or use patterns (e.g. consumption of predatory fish that accumulate high levels of methylmercury), and control at source of production (e.g. emissions of dioxins).

7.10.2.2 Cost–benefit and risk–benefit analyses

While health risk management decisions should be based on risk assessments, a number of other factors will influence 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 aflatoxins versus the cost of implementation of a management strategy) and/or risk–benefit analysis (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 risks and will need to reflect wider societal factors.

7.10.2.3 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 uncertainty factor of 100 when based on animal studies, has been accepted by international institutions and countries as a health-based guidance value. Although 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.

A predicted risk level, such as 10^{-6}, determined from dose–response analysis has been used by some countries and institutions as being not appreciable or negligible (virtually safe dose). 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, 2004b). The United States Occupational Safety and Health Administration has considered a lifetime cancer risk for workers higher than 10^{-3} to represent an unacceptably high risk, and its goal is to reduce this risk to less than 10^{-5} (OSHA, 1983, 1984).

Proposals for the application of lifetime risk estimates in establishing tolerable risk levels have also been published in Europe (Bos et al., 2004).

8. CONCLUSIONS AND RECOMMENDATIONS

8.1 Conclusions

- Full DRM can be considered a more sophisticated or robust alternative to the NOAEL approach in all cases where suitable dose–response data are available (e.g. several dose groups with different response levels).

- For quantal dose–response data, the interest is often in low response (incidence) levels. This may call for low-dose extrapolation by several orders of magnitude (e.g. for tumour incidences). However, equally plausible dose–response risk models may result in highly divergent low estimates. A currently applied approach is to estimate a BMD_{10} and linearly extrapolate from that point downwards, as a conservative approach. Another option, currently under development, is to apply a Bayesian approach that considers the various models all together.

- For continuous dose–response data, two approaches of DRM exist. One is to transform the continuous data into quantal data. The other is to consider continuous dose–response data as information on the severity of the effect and therefore as a function of dose. In the latter approach, measurable changes of effect are often close to response levels considered as adverse (e.g. 10% inhibition of cholinesterase), and the low-dose extrapolation problem is minor or non-existent.

- For the purpose of deriving an ADI, TDI, or RfD, DRM may be used for deriving a BMD, to be used as a point of departure in the same way as the NOAEL is used (i.e. the same uncertainty factors would be applied to the BMD as to the NOAEL).

- DRM may also be used for estimating risks at a given (human) exposure level. For risks in terms of incidences (quantal data), this may involve low-dose extrapolation.

- DRM exercises can provide information on uncertainties associated with the data and identify factors contributing to uncertainties in risk estimates.

- Application of DRM for all end-points can be cost prohibitive, so it is efficient to pre-select the apparently more sensitive end-points. In some

cases, however, it is not easy to identify the most sensitive end-points by visual inspection, so all of the end-points may need to be modelled.

- The BMD and the BMDL should always be reported, so that the quality of the data and the model fit are clear and potencies can be compared on the basis of the BMD.

- The output of the different models used in DRM should be presented.

8.2 Recommendations

- Toxicity testing protocols (e.g. Organisation for Economic Co-operation and Development guidelines) should be reviewed for optimization for BMD and other DRM approaches, including optimal designs for the number of animals and number of doses for different dose–response curves. Additional research is needed for the development of optimal study designs. Guidance should be developed for combining existing studies with a view to DRM.

- Better guidance needs to be developed for combined analysis of different data sets for more precisely estimating BMDs.

- Better understanding of when and how to use the BMR needs to be developed.

- Better understanding of the shape of the dose–response curve at low doses needs to be developed. Additional research is needed to determine the biological basis for extrapolation (e.g. by using biomarkers, tumour precursors, genetically modified animals, and toxicokinetics for target dose estimation).

- Improved guidance needs to be developed for risk communication based on the results of DRM and probabilistic assessment techniques. This should include communication of the types of uncertainty and the relation to statistical variability, imprecision, and the use of confidence intervals.

- The use of DRM should be reviewed and additional general principles for its use developed when more experience becomes available.

REFERENCES

Akaike H (1973) Information theory and an extension of the maximum likelihood principle. In: Petrov BN & Csaki F ed. Proceedings of the second international symposium on information theory. Budapest, Akademiai Kiado, pp 267–281.

Allen BC, Kavlock RJ, Kimmel CA, & Faustman EM (1994) Dose–response assessment for developmental toxicity: II. Comparison of generic benchmark dose estimates with NOAELs. Fundam 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. Fundam Appl Toxicol, **32**: 194–204.

Barlow S, Renwick AG, Kleiner J, Bridges JW, Busk L, Dybing E, Edler L, Eisenbrand G, Fink-Gremmels J, Knaap A, Kroes R, Liem D, Müller DJG, Page S, Schlatter J, Tritscher A, Rolland V, Tueting W, & Würtzen, G (2006) Risk assessment of substances that are both genotoxic and carcinogenic. Food Chem Toxicol, **44**(10): 1636–1650.

Barnes DG & Dourson ML (1988) Reference dose (RfD): Description and use in health risk assessments. Regul Toxicol Pharmacol, **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. Regul Toxicol Pharmacol, **21**: 296–306.

Bos PMJ, Baars BJ, & van Raaij MTM (2004) Risk assessment of peak exposure to genotoxic carcinogens: A pragmatic approach. Toxicol Lett, **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 & Anderson DR (2002) Model selection and multimodel inference, 2nd ed. New York, Springer.

Carrington CD & Bolger PM (2000) A pooled analysis of the Iraqi and Seychelles methylmercury studies. Hum Ecol Risk Assess, **6**: 323–340.

Carrington CD, Bolger PM, & Scheuplein RJ (1996) Risk analysis of dietary lead exposure. Food Addit Contam, **13**: 61–76.

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): 1269–1274 (http://ehp.niehs.nih.gov/docs/2004/6950/abstract.html).

Committee on Carcinogenicity (2003) Minutes of United Kingdom 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).

Crump KS (1984) A new method for determining allowable daily intakes. Fundam Appl Toxicol, **4**: 854–871.

Crump KS (1995) Calculation of benchmark doses from continuous data. Risk Anal, **15**(1): 79–89.

Crump K (2002) Critical issues in benchmark calculations from continuous data. Crit Rev Toxicol, **32**(3): 133–153.

Davidian M & Giltinan DM (1995) Nonlinear models for repeated measurement data. London, Chapman and Hall.

Dourson ML & Stara JF (1983) Regulatory history and experimental support of uncertainty (safety) factors. Regul Toxicol Pharmacol, **2**: 224–238.

Dourson ML, Hertzberg RC, Hartung R, & Blackburn K (1985) Novel approaches for the estimation of acceptable daily intake. Toxicol Ind Health, **1**: 23–41.

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.

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 Chem Toxicol, **40**: 283–326.

Efron B (1987) Better bootstrap confidence intervals. J Am Stat Soc, **82**: 171–185.

Efron B & Tibshirani R (1993) An introduction to the bootstrap. London, Chapman and Hall.

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. Brussels, European Commission, Working Group on the Classification and Labelling of Dangerous Substances.

European Commission (2000) White paper on food safety. Commission of the European Communities, 12 January 2000, Comm (1999) 719 final (http://ec.europa.eu/dgs/health_consumer/library/pub/pub06_en.pdf).

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, & Peacock B (1993) Statistical distributions, 2nd ed. New York, John Wiley & Sons.

Fitzgerald DJ, Robinson NI, & 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**: 1341–1346.

Fowles JR, Alexeeff GV, & Dodge D (1999) The use of benchmark dose methodology with acute inhalation lethality data. Regul Toxicol Pharmacol, **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 (1989) Quantitative risk analysis for quantal reproductive and developmental effects. Environ Health Perspect, **79**: 243–246.

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). Regul Toxicol Pharmacol, **40**: 9–17.

Gaylor D & Slikker W (1994) Modelling for risk assessment of neurotoxic effects. Risk Anal, **14**(3): 333–338.

Gaylor D & Slikker W (2004) Role of the standard deviation in the estimation of benchmark doses with continuous data. Risk Anal, **6**: 1683–1687.

Gaylor DW, Axelrad JA, Brown RP, Cavagnaro JA, Cyr WH, Hulebak KI, Lorentzen RJ, Miller MA, Mulligan LT, & Schwetz BA (1997) Health risk assessment practices in the U.S. Food and Drug Administration. Regul Toxicol Pharmacol, **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. Regul Toxicol Pharmacol, **29**: 151–157.

Gelman A, Carlin JB, Stern HS, & Rubin DB (2004) Bayesian data analysis, 2nd ed. Boca Raton, Florida, Chapman and Hall/CRC.

Ghani AC, Ferguson NM, Donnelly CA, & Anderson RM (2000) Predicted vCJD mortality in Great Britain. Nature, **406**: 583–584.

Gold LS, Gaylor DW, & Slone TH (2003) Comparison of cancer risk estimates based on a variety of risk assessment methodologies. Regul Toxicol Pharmacol, **237**: 45–53.

Hacking I (1976) The great decision. In: The emergence of probability. A philosophical study of early ideas about probability, induction and statistical inference. Cambridge, Cambridge University Press, pp 63–72.

Hasselblad V & Jarabek AM (1995) Dose–response analysis of toxic chemicals. In: Berry DA & Stangl DK ed. Bayesian biostatistics. New York, Marcel Dekker.

Health Canada (1994) Human health risk assessment for priority substances. Ottawa, Ontario, Minister of Supply and Services Canada (http://www.hc-sc.gc.ca/ewh-semt/pubs/contaminants/existsub/approach/index_e.html).

Hertzberg RC (1991) Quantitative extrapolation of toxicological finds. In: Krewski D & Franklin C ed. Statistics in toxicology. New York, Gordon & Breach Science.

Hertzberg RC & Miller M (1985) A statistical model for species extrapolation using categorical response data. Toxicol Ind Health, **1**(4): 43–57.

Hertzberg RC & Wymer L (1991) Modelling the severity of toxic effects. In: Proceedings of the 84th annual meeting and exhibition of the Air and Waste Management Association, 16–21 June 1991, Vancouver, British Columbia. Pittsburgh, Pennsylvania, Air and Waste Management Association.

Humphreys SH, Carrington CD, & Bolger PM (2001) A quantitative risk assessment for fumonisins B_1 and B_2 in corn. Food Addit Contam, **18**: 211–220.

IPCS (1987) Principles for the safety assessment of food additives and contaminants in food. Geneva, World Health Organization, International Programme on Chemical Safety (Environmental Health Criteria 70; http://www.inchem.org/documents/ehc/ehc/ehc70.htm).

IPCS (1994) Assessing human health risks of chemicals: Derivation of guidance values for health-based exposure limits. Geneva, World Health Organization, International Programme on Chemical Safety (Environmental Health Criteria 170; http://www.inchem.org/documents/ehc/ehc/ehc170.htm).

IPCS (1999) Principles for the assessment of risks to human health from exposure to chemicals. Geneva, World Health Organization, International Programme on Chemical Safety (Environmental Health Criteria 210; http://www.inchem.org/documents/ehc/ehc/ehc210.htm).

IPCS (2000) Human exposure assessment. Geneva, World Health Organization, International Programme on Chemical Safety (Environmental Health Criteria 214; http://www.inchem.org/documents/ehc/ehc/ehc214.htm).

IPCS (2004) IPCS risk assessment terminology. Geneva, World Health Organization, International Programme on Chemical Safety (Harmonization Project Document No. 1; http://www.who.int/ipcs/methods/harmonization/areas/ipcsterminologyparts1and2.pdf).

IPCS (2005) Chemical-specific adjustment factors for interspecies differences and human variability: Guidance document for use of data in dose/concentration–response assessment. Geneva, World Health Organization, International Programme on Chemical Safety (Harmonization Project Document No. 2; http://whqlibdoc.who.int/publications/2005/9241546786_eng.pdf).

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**: 287–296.

Kimmel CA & Gaylor DW (1988) Issues in qualitative and quantitative risk analysis for development toxicology. Risk Anal, **8**: 15–20.

Kodell RL, Chen JJ, & Gaylor DW (1995) Neurotoxicity modeling for risk assessment. Regul Toxicol Pharmacol, **22**(1): 24–29.

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 Chem Toxicol, **40**: 327–385.

Lehman AJ & Fitzhugh OG (1954) Ten-fold safety factor studies. Assoc Food Drug Off US Q Bull, **XVIII**(1): 33–35.

Moerbeek M, Piersma AH, & Slob W (2004) A comparison of three methods for calculating confidence intervals for the benchmark dose. Risk Anal, **24**(1): 31–40.

Murrell JA, Portier CJ, & Morris RW (1998) Characterizing dose–response. I. Critical assessment of the benchmark dose concept. Risk Anal, **18**: 13–26.

National Health and Medical Research Council (1999) Toxicity assessment for carcinogenic soil contaminants. Canberra, Commonwealth of Australia.

NRC (1983) Risk assessment in the federal government: Managing the process. Washington, DC, National Research Council, National Academy of Science Press.

NRC (1994) Science and judgement in risk assessment. Washington, DC, National Research Council, National Academy of Science Press.

NRC (1996) Understanding risk: Informing decisions in a democratic society. Washington, DC, National Research Council, National Academy of Science Press.

NTP (2002) Report on carcinogens, 10th ed. Research Triangle Park, North Carolina, United States Department of Health and Human Services, National Institutes of Health, National Toxicology Program (http://ntp-server.niehs.nih.gov/index.cfm?objectid=72016262-BDB7-CEBA-FA60E922B18C2540).

O'Brien J, Renwick AG, Constable A, Dybing E, Müller DJG, Schlatter J, Slob W, Tueting W, van Benthem J, Williams GM, & Wolfreys A (2006) Approaches to the risk assessment of genotoxic carcinogens in food: A critical appraisal. Food Chem Toxicol, **44**(10): 1613–1635.

OSHA (1983) Occupational exposure to inorganic arsenic: Supplemental statement of reasons for the final rule, 48 F.R. 1,864. Washington, DC, United States Department of Labor, Occupational Safety and Health Administration.

OSHA (1984) Occupational exposure to ethylene oxide: Final standard, 49 F.R. 46,936. Washington, DC, United States Department of Labor, Occupational Safety and Health Administration.

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. Reprod Toxicol, **14**: 417–425.

Pieters MN, Bakker M, & Slob W (2004) Reduced intake of deoxynivalenol in the Netherlands: A risk assessment update. Toxicol Lett, **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 Compds, **29**: 222–227.

Presidential Commission (1997) Risk assessment and risk management in regulatory decision-making, Vol 2. Washington, DC, The Presidential/Congressional Commission on Risk Assessment and Risk Management.

Raftery A, Madigan D, & Hoeting J (1997) Bayesian model averaging for linear regression models. J Am Stat Assoc, **92**: 179–191.

Renwick AG & Lazarus NR (1998) Human variability and noncancer risk assessment—an analysis of the default uncertainty factor. Regul Toxicol Pharmacol, **27**: 3–20.

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, & Kroes R (2003) Risk characterization of chemicals in food and diet. Food Chem Toxicol, **41**: 1211–1271.

Rescher N (1969) Many-valued logic. New York, McGraw-Hill.

Sand S, Falk Filipsson A, & Victorin K (2002) Evaluation of the benchmark dose method for dichotomous data: Model dependence and model selection. Regul Toxicol Pharmacol, **36**: 184–197.

Sand S, von Rosen D, & Falk Filipsson A (2003) Benchmark calculations in risk assessment using continuous dose response information: The influence of variance and the determination of a cut-off value. Risk Anal, **23**: 1059–1068.

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, **88**: 331–341.

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

Sherman CD & Portier CJ (1998) Multistage carcinogenesis models. In: Encyclopedia of biostatistics. Sussex, Wiley & Sons, pp 2808–2814.

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

Slob W, Moerbeck M, Rauniomaa E, & Piersma AH (2005) A statistical evaluation of toxicity designs for the estimation of the benchmark dose in continuous endpoints. Toxicol Sci, **84**: 167–185.

Thompson KM (2002) Variability and uncertainty meet risk management and risk communication. Risk Anal, **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. Toxicol Sci, **66**: 34–53.

USEPA (1986) Guidelines for carcinogen risk assessment. United States Environmental Protection Agency. Fed Regist, **51**: 33992–34003.

USEPA (1996) Proposed guidelines for carcinogen risk assessment. United States Environmental Protection Agency. Fed Regist, **61**: 17960–180011.

USEPA (2005) Guidelines for carcinogen risk assessment 2005. Washington, DC, United States Environmental Protection Agency, Risk Assessment Forum (EPA/630/P-03/001F; http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=116283).

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) Evaluation of certain food additives and contaminants. Forty-ninth report of the Joint FAO/WHO Expert Committee on Food Additives. Geneva, World Health Organization.

WHO (2000) Air quality guidelines for Europe, 2nd ed. Copenhagen, World Health Organization Regional Office for Europe.

WHO (2002) Evaluation of certain food additives and contaminants. Fifty-seventh report of the Joint FAO/WHO Expert Committee on Food Additives. Geneva, World Health Organization (Technical Report Series 909).

WHO (2003) Health aspects of air pollution with particulate matter, ozone and nitrogen dioxide. Report of a Working Group. Copenhagen, World Health Organization Regional Office for Europe (http://www.euro.who.int/document/e79097.pdf).

WHO (2004a) Risk assessment of *Listeria monocytogenes* in ready-to-eat foods. Geneva, World Health Organization (Microbiological Risk Assessment Series, No. 5, Technical Report; http://www.who.int/foodsafety/publications/micro/mra_listeria/en/index.html).

WHO (2004b) Guidelines for drinking-water quality, 3rd ed. Vol. 1. Recommendations. Geneva, World Health Organization (http://www.who.int/water_sanitation_health/dwq/gdwq3rev/en/).

WHO (2006) Evaluation of certain food contaminants. Sixty-fourth report of the Joint FAO/WHO Expert Committee on Food Additives. Geneva, World Health Organization (WHO Technical Report Series No. 930; http://whqlibdoc.who.int/trs/WHO_TRS_930_eng.pdf).

WHO (2007a) Evaluation of certain food additives and contaminants. Sixty-seventh report of the Joint FAO/WHO Expert Committee on Food Additives. Geneva, World Health Organization (WHO Technical Report Series No. 940; http://whqlibdoc.who.int/trs/WHO_TRS_940_eng.pdf).

WHO (2007b) Evaluation of certain food additives and contaminants. Sixty-eighth report of the Joint FAO/WHO Expert Committee on Food Additives. Geneva, World Health Organization (WHO Technical Report Series No. 947; http://whqlibdoc.who.int/trs/WHO_TRS_947_eng.pdf).

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

Woutersen RA, Jonker D, Stevenson H, te Biesebeek JD, & 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 Chem Toxicol, **39**: 697–707.

ANNEX 1: TERMINOLOGY

Acceptable daily intake (ADI)/tolerable daily intake (TDI)/reference dose (RfD): 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 the individual's lifetime without appreciable health risk.

Acceptable risk: A risk management term. The acceptability of risk depends on scientific data, on 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 lifespan 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 criterion: 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 exposure to an agent below which an adverse effect is not likely to occur (see *Safety factor* and *Uncertainty factor*).

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): A dose of a substance associated with a specified low incidence of risk, generally in the range of 1–10%, 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 a 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 $0 < p < 1$.

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 resampling with replacement of the sample values or resampling 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 weight per 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 (CSAF): 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 (one-sided): An interval below the estimated upper confidence limit, or an interval above the estimated lower confidence limit, that is expected to include the true value of an estimated parameter with a specified confidence (percentage of the time).

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

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.

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.

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 or taken up or absorbed by an organism, system, or (sub)population.

Dose–response: Relationship between the amount of an agent administered to, taken up by, 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 by, 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_x: Effective dose associated with a biological effect in x% of the individuals. Dose may be the external exposure often expressed in milligrams of the substance per day per kilogram body weight raised to a power (generally 1, 3/4, or 2/3) or area under the curve (AUC) in blood or target tissue 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.

Gamma distribution: A unimodal statistical distribution (relative proportion of responders as a function of some measure) 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.

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 an 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 percentage 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 function: Relative probabilities that various values of population parameters would arise from the sample observations.

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.

Logistic model: A sigmoidal (S-shaped) function that relates the proportion of individuals with a specified characteristic to an independent variable.

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 dose of a substance, found by experiment or observation, that causes an adverse alteration of morphology, functional capacity, growth, development, or lifespan 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 dose of a substance, found by experiment or observation, that causes any alteration of morphology, functional capacity, growth, development, or lifespan of the target organisms distinguishable from normal (control) organisms of the same species and strain under the same defined conditions of exposure.

Margin of exposure (MOE): Ratio of the no-observed-adverse-effect level (NOAEL) or benchmark dose lower confidence limit (BMDL) for the critical effect to the theoretical, predicted, or estimated exposure dose or concentration.

Maximum likelihood estimate: 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.

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 benchmark dose.

Non-linear 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.

No-observed-adverse-effect level (NOAEL): The highest concentration or dose of a substance, found by experiment or observation, that causes no detectable adverse alteration of morphology, functional capacity, growth, development, or lifespan of the target organisms under defined conditions of exposure.

No-observed-effect level (NOEL): The highest concentration or dose of a substance, found by experiment or observation, that causes no detectable alteration of morphology, functional capacity, growth, development, or lifespan of the target organisms under defined conditions of exposure.

Normal distribution: A mathematical description where a continuous random variable x with a mean μ and a variance σ^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.

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: 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. 1–10% incidence of an effect). Depending on the mode of action and available data, some form of extrapolation below the point of departure may be employed for low-dose risk assessment, or the point of departure may be divided by a series of uncertainty factors to arrive at a reference dose. Points of departure include the BMD, BMDL, LOAEL, and carcinogenic potency estimates, such as the T_{25}.

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 sigmoidal shape.

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

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.

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.

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

Repeated measures: A biological end-point 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.

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 *Assessment factor* and *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 which 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 *Assessment factor* and *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 levelling 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.

RESUME, CONCLUSIONS ET RECOMMANDATIONS

1. Résumé

La modélisation de la relation dose-réponse (DRM), destinée à évaluer quantitativement les risques et enfin à étayer les décisions de santé publique concernant les expositions à des produits chimiques, peut être décrite comme un processus en six étapes. Les quatre premières étapes – sélection des données, choix du modèle, mise en relation statistique et estimation des paramètres – constituent une analyse de la relation dose-réponse. Ces étapes composent le processus permettant d'obtenir une description mathématique des données, en vue de prédire les réponses à des doses connues ou d'établir des estimations de doses à partir d'une réponse donnée. La cinquième étape réalise une synthèse des résultats de l'analyse dose-réponse et des estimations de l'exposition afin de guider les décisions de santé publique. L'étape finale, qui peut éventuellement intervenir plus tôt dans la DRM, évalue la qualité de la relation dose-réponse et la sensibilité des prédictions aux hypothèses ayant servi à l'analyse.

La caractérisation de la relation dose-réponse dans les études chez l'homme et chez l'animal constitue une composante majeure de la caractérisation des dangers et sert à extrapoler les incidences des effets nocifs sur la gamme de niveaux d'exposition humaine. Au cours des années, diverses méthodes ont été mises au point pour traiter de telles relations, améliorer l'extrapolation pour les faibles doses et dériver des valeurs guides reposant sur des considérations sanitaires telles que les doses journalières admissibles (DJA), les doses journalières tolérables (DJT) et les doses de référence (D_{ref}). La DRM peut s'avérer utile dans les évaluations des risques en permettant un meilleur usage des données disponibles et en fournissant des outils pour évaluer la qualité des données et les incertitudes résultantes sur les estimations de la relation dose-réponse.

D'une manière générale, les estimations obtenues par DRM sont établies à partir de données provenant de l'ensemble de la courbe dose-réponse pour l'effet critique. L'approche standard reposant sur la dose sans effet nocif observé (DSENO) peut être considérée comme un cas spécial et simplifié d'analyse de la

relation dose-réponse, dans la mesure où elle identifie une dose unique supposée ne pas avoir d'effet nocif appréciable. La modèle DRM reflète les caractéristiques de la courbe dose-réponse, en permettant notamment d'estimer sa pente. Dans le cas d'un cadre de régression, il indique l'écart-type et l'intervalle de confiance pour les paramètres de modélisation. L'un des inconvénients de l'approche DSENO réside dans l'impossibilité de quantifier les degrés de variabilité et d'incertitude, alors que d'autres modèles de la relation dose-réponse peuvent faciliter l'analyse de ces grandeurs. L'utilisation d'un modèle dose-réponse peut permettre d'optimiser la conception de l'étude et de préciser les besoins en matière d'études supplémentaires. L'approche DSENO intègre des informations biologiques à travers l'application d'un jugement d'expert, néanmoins subjectif. Une DRM complète est en mesure de fournir une analyse plus « riche en éléments scientifiques » grâce à l'inclusion quantitative plus formelle dans les modèles de facteurs et de covariables, par exemple. Les estimations obtenues à partir de cette DRM facilitent la comparaison dans un cadre commun entre des expériences, des effets et des composés qui diffèrent sur le plan quantitatif. La modélisation DRM peut aussi conduire à de meilleures évaluations des risques et de l'innocuité et offre la possibilité d'étudier les probabilités d'effets se manifestant en dehors de la plage observable.

Le choix des modèles à utiliser dépend du type de donnée. Il faut sélectionner un modèle dose-réponse et un modèle décrivant la variabilité des données. Une fois qu'on dispose de modèles adaptés à un jeu de données, on peut évaluer leur degré de représentativité pour ces données par des mesures de la qualité de l'ajustement. On peut en outre comparer leur capacité à décrire les données. Les incertitudes portant sur les inférences tirées de ces modèles se répartissent dans quatre catégories : incertitudes statistiques sur les inférences dues à la variabilité des réponses entre les sujets des expériences, erreurs expérimentales (randomisation imparfaite, erreurs de dosage, localisation défavorable de la dose, par exemple), variabilité d'une expérience à l'autre due aux différences inévitables dans l'exécution de l'expérience et incertitude due au fait que l'on ne connaît pas le « vrai modèle » décrivant les données. L'analyse de la relation dose-réponse doit, dans la mesure du possible, prendre en compte l'ensemble de ces quatre sources de variabilité et d'incertitude.

Le calcul des doses de référence (BMD) est une application particulièrement importante de la DRM. Les BMD sont les doses pour lesquelles on détermine par déduction qu'il se produira un niveau donnée de réponse. Lorsqu'on dispose de données appropriées, les BMD offrent une alternative à l'approche DSENO pour le calcul de valeurs guides reposant sur des considérations sanitaires. Lorsqu'une extrapolation s'avère nécessaire, il convient de représenter l'incertitude associée à la prédiction. Il est dans ce cas particulièrement important d'indiquer l'incertitude liée au modèle.

Une DRM complète peut apporter des informations supplémentaires au gestionnaire de risques. La sortie du modèle doit être conçue pour répondre à certaines questions concernant la probabilité d'effets sanitaires nocifs. Elle peut être présentée essentiellement de trois façons. Premièrement, elle peut servir à établir des valeurs guides reposant sur des considérations sanitaires telles que les DJA, les DJT ou les D_{ref}, d'une manière analogue aux procédures actuellement appliquées à partir de la DSENO ou de la dose minimale avec effets nocifs observés (DMENO). La DRM peut être une méthode plus sûre sur le plan scientifique pour déterminer ces valeurs guides. Deuxièmement, la sortie de la DRM peut être utilisée en gestion des risques pour estimer une marge d'exposition (ME), par détermination du rapport de la dose correspondant à une limite donnée de la réponse à un niveau d'exposition humaine. Troisièmement, sur la base de la relation dose-réponse modélisée, cette sortie peut être une estimation quantitative de l'ampleur du risque ou de l'effet sanitaire pour un niveau d'exposition humaine, moyennant l'hypothèse généralement acceptée que les facteurs d'incertitude utilisés couvrent les incertitudes associées aux différences de sensibilité entre individus et espèces. La DRM peut fournir de meilleures informations sur la probabilité des effets pour les doses faibles et inférieures aux niveaux observés dans les systèmes biologiques, ainsi que de meilleures estimations des incertitudes statistiques entachant les estimations des effets probables.

La multiplicité des jeux de données et des incertitudes peut influer sur le type de sortie des exercices de DRM et avoir de l'importance pour les gestionnaires de risques. On peut utiliser la DRM avec des données d'exposition pour identifier les sous-

populations à risque. Elle peut aussi aider les gestionnaires dans la détermination des priorités et dans l'évaluation des conséquences d'interventions proposées pour réduire les risques. Pour la communication à propos des risques, l'application des techniques de DRM offre des opportunités, mais comporte aussi des difficultés. Les évaluations par DRM peuvent produire des informations sous plusieurs formats, et notamment sous forme de fonctions dose-réponse permettant, avec les estimations de l'exposition, de prédire les risques pour des niveaux d'exposition donnés, ainsi que de fonctions permettant inversement d'estimer les niveaux d'exposition à l'origine de risques donnés. On obtient ainsi notamment des estimations du risque potentiel d'absorption plus importante qu'une valeur guide reposant sur des considérations sanitaires, DJA par exemple. Ces évaluations offrent aussi des approches pour comparer les risques ou les bénéfices concurrents et s'intéresser aux incertitudes susceptibles d'influer sur les risques prédits. Toutefois, à moins que la situation en termes de risque ne soit envisagée à l'échelle de la population, il existe un problème de communication à propos du risque car lorsqu'on présente le niveau de risque dans des situations où aucun niveau d'exposition n'est dépourvu de risque, la modélisation prévoit qu'un certain pourcentage de la population subira des effets jugés nocifs. Il faut reconnaître que l'utilisation de la DRM impose aux données des exigences en termes de qualité et de quantité et nécessite des compétences spécifiques.

L'utilisation courante des estimations tirées der la DRM pourrait, du point de vue de la gestion des risques, permettre une meilleure caractérisation avant la prise de décisions en :

- apportant des informations sur les valeurs guides (ampleur et types des impacts sanitaires) ;
- montrant les bénéfices de différentes actions réglementaires ;
- fournissant au décideur une appréciation des données « plus que ponctuelle » ;
- favorisant la cohérence dans les décisions, moyennant des ajustements appropriés pour tenir compte des différences entre les effets, les niveaux d'effet, les espèces et les types d'étude ; et en
- permettant en continu et en permanence des interactions itératives entre l'évaluateur et le gestionnaire de risques.

L'utilisation de la DRM et des techniques d'évaluation probabiliste pour décrire quantitativement la variabilité et l'incertitude génère de nouvelles difficultés dans la communication à propos des risques. Ces difficultés résident notamment dans :

- l'explication de la prévision, pour un certain pourcentage de la population, d'un dépassement du niveau de sécurité et/ou de l'apparition d'effets nocifs ;
- l'explication du niveau de risque dans les cas où on suppose qu'aucun niveau d'exposition n'est dépourvu de risque ;
- la comparaison entre risques ou bénéfices concurrents ;
- la mise en lumière d'incertitudes influant sur le risque prédit ; et
- l'explication du fait qu'en matière de risque, une estimation indique ce qui peut se passer au niveau d'une population, plutôt qu'à celui d'un individu, ce qui, notons le, vaut aussi pour l'approche DJA/DJT.

2. Conclusions

- La DRM complète peut être considérée comme une alternative plus élaborée et plus robuste à l'approche DSENO dans tous les cas où l'on dispose de données appropriées sur la relation dose-réponse (pour plusieurs groupes de dose et différents niveaux d'exposition, par exemple).

- Pour les données dose-réponse ponctuelles, on s'intéresse souvent aux faibles niveaux de réponse (d'incidence). Il est parfois nécessaire, dans cette perspective, d'extrapoler sur plusieurs ordres de grandeur (pour l'incidence des tumeurs, par exemple). Cependant, des modèles également plausibles de la relation dose-réponse peuvent fournir des estimations fortement divergentes pour les faibles valeurs. Une approche actuellement appliquée en tant que méthode prudente consiste à estimer la BMD_{10} (dose pour un risque de 10 %) et à extrapoler linéairement à partir de ce point vers les valeurs descendantes. Une autre solution, actuellement en cours de développement, applique une approche bayésienne, considérant globalement les divers modèles.

- Pour les données dose-réponse continues, il existe deux approches de type DRM. L'une comprend la transformation des

données continues en données ponctuelles. L'autre considère les données dose-réponse continues comme des informations sur la gravité de l'effet et donc comme une fonction de la dose. Dans cette dernière approche, des variations mesurables de l'effet sont souvent proches des niveaux de réponses considérés comme nocifs (par exemple, inhibition de 10 % de la cholinestérase) et le problème de l'extrapolation à faible dose est mineur ou ne se pose pas.

- Pour dériver une DJA, une DJT ou une Dref, on peut faire appel à la DRM pour déterminer une BMD, qui sera utilisée comme point de départ de la même façon qu'une DSENO (c'est-à-dire qu'on appliquera les mêmes facteurs d'incertitude à la BMD qu'à la DSENO).

- La DRM peut aussi être employée pour estimer les risques correspondant à un niveau d'exposition (humaine) donné. Pour évaluer les risques en termes d'incidence (données ponctuelles), cette opération peut devoir inclure une extrapolation aux faibles doses.

- Les exercices de DRM peuvent apporter des informations sur les incertitudes associées aux données et identifier des facteurs contribuant aux incertitudes sur les estimations des risques.

- L'application de la DRM à tous les points finaux peut être extrêmement onéreuse, il est donc plus efficace de présélectionner les points finaux apparemment les plus sensibles. Dans certains cas cependant, il n'est pas facile d'identifier visuellement ces points de sorte qu'il peut être nécessaire de modéliser tous les points finaux.

- La BMD et la borne inférieure de l'intervalle de confiance de la BMD (BMDL) doivent toujours être indiquées de manière à ce que la qualité des données et de l'ajustement du modèle apparaisse clairement et que l'on puisse procéder à une comparaison de puissances à partir de la BMD.

- Il convient de présenter la sortie des différents modèles de DRM.

3. Recommandations

- Les protocoles d'évaluation de la toxicité (par exemple les Lignes directrices de l'Organisation de coopération et de développement économiques) doivent être examinés pour optimiser l'approche utilisant la BMD et d'autres approches de type DRM, notamment pour choisir au mieux les nombres d'animaux et de doses pour les différentes courbes dose-réponse. Des recherches supplémentaires sont nécessaires pour développer des types d'étude optimaux. Il faut également élaborer des conseils pour combiner les études existantes en vue d'une DRM.

- Il faut mettre au point des recommandations pour l'analyse combinée de différents jeux de données en vue d'une estimation plus précise des BMD.

- Il faut aussi parvenir à mieux comprendre quand et comment utiliser la réponse de référence (BMR).

- La forme de la courbe dose-réponse aux faibles doses doit être mieux interprétée. Des recherches supplémentaires sont nécessaires pour déterminer la base biologique de l'extrapolation (en faisant appel, par exemple, à des marqueurs biologiques, à des précurseurs de tumeur, à des animaux génétiquement modifiés ou à la toxico-cinétique, pour estimer la dose cible).

- Il faut élaborer de meilleures recommandations pour la communication à propos des risques sur la base des résultats de la DRM et des techniques d'évaluation probabiliste. Cette communication devra couvrir les types d'incertitude, leur relation avec la variabilité statistique, l'imprécision et l'utilisation des intervalles de confiance.

- L'utilisation de la DRM doit faire l'objet d'un bilan et des principes généraux supplémentaires devront être développés à mesure que l'on disposera de plus d'expérience.

RESUMEN, CONCLUSIONES Y RECOMENDACIONES

1. Resumen

La creación de modelos de la relación dosis-respuesta, para su utilización en la evaluación cuantitativa del riesgo y en último término para documentar las decisiones en materia de salud pública, se puede describir como un proceso en seis etapas. Las cuatro primeras etapas—selección de datos, selección del modelo, vinculación estadística y estimación de los parámetros—constituyen el análisis de la relación dosis-respuesta. Estas etapas están relacionadas con el proceso mediante el cual se obtiene una descripción matemática de los datos, a fin de evaluar respuestas previstas para dosis conocidas u obtener estimaciones de la dosis cuando lo que interesa es una respuesta determinada. La quinta etapa consiste en la integración de los resultados del análisis de la relación dosis-respuesta en las estimaciones de la exposición, con el objetivo de orientar las decisiones relativas a la salud pública. La última etapa, que se puede elegir aplicar antes, consiste en una evaluación de la calidad del análisis de la relación dosis-respuesta y de la sensibilidad de las predicciones de los modelos con respecto a las hipótesis utilizadas en el análisis.

La caracterización de las relaciones dosis-respuesta en estudios realizados en animales y personas ha sido un componente importante de la caracterización del peligro y se ha utilizado en la extrapolación de incidencias de efectos adversos en la gama de los niveles de exposición humana. Durante años se han elaborado diversos métodos para ajustar dichas relaciones, mejorar la extrapolación a dosis bajas y obtener valores guía basados en la salud, como la ingesta diaria admisible (IDA), la ingesta diaria tolerable (IDT) y las dosis de referencia. La creación de modelos puede ser útil en las evaluaciones del riesgo para utilizar mejor los datos disponibles y para suministrar instrumentos de evaluación de la calidad de los datos y las consiguientes incertidumbres en las estimaciones de la relación dosis-respuesta.

En general, las estimaciones de los modelos de la relación dosis-respuesta se basan en los datos obtenidos de la totalidad de la

curva correspondiente a dicha relación para el efecto crítico. El método normalizado de la concentración sin efectos adversos observados (NOAEL) se puede considerar como un caso especial simplificado de análisis de la relación dosis-respuesta, puesto que identifica una dosis única que se supone que no tiene un efecto adverso apreciable. El modelo de la relación dosis-respuesta refleja las características de la curva de dicha relación, en particular porque proporciona estimaciones de la pendiente. En el caso de un marco de regresión, proporciona el error estándar y los intervalos de confianza para los parámetros del modelo. La utilización del método de la NOAEL tiene el inconveniente de que no es posible cuantificar el grado de variabilidad e incertidumbre que puede haber, mientras que otros modelos de la relación dosis-respuesta pueden facilitar el análisis de la sensibilidad y la incertidumbre. El examen de un modelo de dosis-respuesta puede mejorar al máximo la formulación del estudio y aclarar la necesidad de estudios adicionales. El método de la NOAEL incorpora información biológica mediante la aplicación de un parecer "experto", pero subjetivo. La creación de modelos de la relación dosis-respuesta completos permitiría un análisis más "científico", por ejemplo mediante la inclusión cuantitativa más oficial de factores y covariantes en los modelos. Las estimaciones derivadas de los modelos de dosis-respuesta mejoran la capacidad para comparar experimentos, efectos y compuestos con diferencias cuantitativas en el ámbito de un marco común. Los modelos pueden mejorar las evaluaciones del riesgo y de la inocuidad, ofreciendo al mismo tiempo oportunidades para examinar la probabilidad de los efectos fuera de la gama observable.

La elección de los modelos que se van a utilizar depende del tipo de datos. Dichos modelos deben incluir un patrón para la relación dosis-respuesta y otro para la variabilidad de los datos. Una vez ajustados los modelos a una serie de datos, se puede evaluar el grado en que los describen individualmente utilizando medidas de la precisión del ajuste. Además, se puede comparar entre ellos la capacidad para describir los datos. Las incertidumbres sobre las consecuencias que puedan derivarse de dichos modelos entran en cuatro categorías principales: incertidumbre estadística de las consecuencias debida a la variabilidad entre las respuestas de los sujetos objeto de experimentación, errores experimentales (por ejemplo, distribución al azar imperfecta, errores de dosificación,

localización desfavorable de las dosis), variabilidad entre experimentos debida a diferencias inevitables en su realización e incertidumbre debida al hecho de que no se conoce el "verdadero modelo" para los datos. Siempre que sea posible, en el análisis de la relación dosis-respuesta hay que abordar las cuatro fuentes de variabilidad e incertidumbre.

Una aplicación particularmente importante de los modelos de dosis-respuesta es el cálculo de las dosis de referencia. Son las dosis con las cuales se deduce que se producirá un determinado nivel de respuesta. Cuando se dispone de datos apropiados, las dosis de referencia son una alternativa al método de la NOAEL para calcular los valores guía basados en la salud. Cuando es necesaria una extrapolación, se debe representar la incertidumbre asociada con una predicción. En este caso es particularmente importante incluir la incertidumbre del modelo.

La creación de modelos de la relación dosis-respuesta completos ofrece la posibilidad de proporcionar información adicional a los gestores del riesgo. Los resultados de los modelos se deben orientar hacia el examen de cuestiones específicas relativas a la probabilidad de efectos adversos en la salud. Se pueden presentar de tres maneras principales. En primer lugar, se pueden utilizar para el establecimiento de un valor guía basado en la salud, por ejemplo una IDA, una IDT o unas dosis de referencia, de manera análoga a los procedimientos actuales basados en la NOAEL o la concentración más baja con efectos adversos observados (LOAEL). Los modelos de dosis-respuesta pueden ser un método más sólido desde el punto de vista científico para determinar valores guía basados en la salud. En segundo lugar, los resultados de dichos modelos se pueden utilizar en la gestión del riesgo para estimar un margen de exposición, mediante el cálculo de la relación entre la dosis correspondiente a un límite determinado de respuesta y un nivel de exposición humana. En tercer lugar, sobre la base de la relación dosis-respuesta obtenida mediante el modelo, el resultado puede ser una estimación cuantitativa de la magnitud del riesgo/efecto en la salud para el nivel de exposición humana, con la hipótesis generalmente aceptada de que los factores de incertidumbre utilizados incluyen las incertidumbres relativas a las diferencias de sensibilidad intraespecíficas e interespecíficas. Los modelos de dosis-respuesta pueden proporcionar mejor información sobre la probabilidad de efectos con dosis bajas, inferiores a los

niveles observados en los sistemas biológicos, y pueden proporcionar asimismo mejores estimaciones de las incertidumbres estadísticas de los efectos probables.

Dos factores que pueden influir en el tipo de resultados obtenidos de la aplicación de los modelos de dosis-respuesta y que pueden ser importantes para el gestor del riesgo son las series de datos múltiples y las incertidumbres. Los modelos se pueden utilizar con datos de exposición para identificar las subpoblaciones en situación de riesgo. También se pueden emplear para ayudar a los gestores del riesgo a establecer prioridades y evaluar las consecuencias de las intervenciones propuestas encaminadas a reducir el riesgo. Para la comunicación del riesgo, la utilización de técnicas con modelos de dosis-respuesta ofrece oportunidades y retos. Las evaluaciones con estos modelos pueden generar información de varios tipos, como funciones de la relación dosis-respuesta que permiten, junto con las estimaciones de la exposición, la predicción de los riesgos con niveles específicos de exposición y funciones que permiten la estimación de los niveles de exposición que dan lugar a riesgos determinados. Esto incluye las estimaciones del posible riesgo de ingestas por encima de un valor guía basado en la salud, por ejemplo la IDA. Las evaluaciones con los modelos de dosis-respuesta también ofrecen métodos para comparar riesgos o beneficios competitivos y permiten concentrar la atención en las incertidumbres que pueden influir en el riesgo pronosticado. Sin embargo, salvo que la situación del riesgo se examine en la población, su comunicación presenta el problema de que, aun explicando el nivel de riesgo en esas circunstancias en las que no hay un nivel inocuo de exposición, cabe predecir que cierto porcentaje de la población va a registrar algunos efectos considerados adversos. Hay que reconocer que la utilización de los modelos de la relación dosis-respuesta requiere cierta cantidad y calidad de datos, así como conocimientos técnicos específicos.

El uso potencial "continuo" de las estimaciones derivadas de los modelos de dosis-respuesta puede, desde una perspectiva de gestión del riesgo, mejorar la caracterización para la adopción de decisiones, porque:

- facilita información sobre lo que ocurre por encima del valor guía basado en la salud (magnitud y tipos de efectos en la salud);
- demuestra los beneficios de distintas medidas normativas;
- ofrece a los encargados de la adopción de decisiones una apreciación de los datos desde más de un punto de vista;
- promueve la coherencia en las decisiones, si se hacen ajustes apropiados para las diferencias en los efectos, el nivel de los efectos, las especies y la formulación del estudio; y
- facilita una interacción iterativa entre el asesor del riesgo y el gestor del riesgo de manera continua e ininterrumpida.

La utilización de modelos de la relación dosis respuesta y de técnicas de evaluación probabilística para describir de manera cuantitativa la variabilidad y la incertidumbre incorpora nuevos retos a la comunicación del riesgo. Algunos de ellos son los siguientes:

- explicar que se prevé que un cierto porcentaje de la población superará el nivel de inocuidad y/o sufrirá un efecto adverso;
- explicar el nivel de riesgo en esas circunstancias en las que se supone que no hay un nivel inocuo de exposición;
- comparar los riesgos o los beneficios en pugna;
- prestar una atención especial a las incertidumbres que influyen en el riesgo pronosticado; y
- explicar que una estimación del riesgo se refiere a lo que puede ocurrir a la población, más que a nivel individual, y señalar que esto es lo que ocurre también con el enfoque de la IDA/IDT.

2. Conclusiones

- La creación de modelos de la relación dosis-respuesta completos se puede considerar un método alternativo más complejo o válido que el de la NOAEL en todos los casos en que se disponga de datos apropiados de la relación dosis-respuesta (por ejemplo, para varios grupos de dosis con distintos niveles de respuesta).

- Para los datos cuantales de la relación dosis-respuesta, el interés radica con frecuencia en los niveles bajos de respuesta (incidencia). Esto puede exigir una extrapolación a dosis más

bajas en varios órdenes de magnitud (por ejemplo, para las incidencias de tumores). Sin embargo, los modelos del riesgo de la relación dosis-respuesta que sean igualmente admisibles pueden dar lugar a estimaciones bajas muy divergentes. Un método aplicado actualmente, considerado prudente, consiste en estimar una dosis de referencia$_{10}$ (dosis con un riesgo del 10%) y hacer una extrapolación de manera lineal descendente desde ese punto. Otra opción, todavía en preparación, consiste en aplicar un método bayesiano, que examina los distintos modelos en conjunto.

- Para la obtención de datos continuos de la relación dosis-respuesta hay dos sistemas de utilización de los modelos. Uno consiste en transformar los datos continuos en datos cuantales. El otro en considerar los datos continuos de la relación dosis-respuesta como información de la gravedad del efecto y, por consiguiente, como una función de la dosis. En el segundo sistema, los cambios mensurables de los efectos suelen estar cerca de los niveles de respuesta considerados adversos (por ejemplo, la inhibición del 10% de la colinesterasa) y el problema de la extrapolación a dosis bajas es insignificante o inexistente.

- Con el fin de obtener un valor de la IDA, la IDT o las dosis de referencia, se pueden utilizar los modelos de dosis-respuesta para derivar una dosis de referencia, que se utilizará como punto de partida de la misma manera que se utiliza la NOAEL (es decir, se aplicarían a la dosis de referencia los mismos factores de incertidumbre que a la NOAEL).

- También se pueden utilizar modelos de la relación dosis-respuesta para estimar los riesgos en un determinado nivel de exposición (humana). Para los riesgos expresados como incidencias (datos cuantales) puede ser necesaria la extrapolación a dosis bajas.

- El uso de modelos de dosis-respuesta puede proporcionar información sobre las incertidumbres asociadas con los datos e identificar los factores que contribuyen a ellas en las estimaciones del riesgo.

- La aplicación de modelos de la relación dosis-respuesta a todos los efectos finales puede tener un costo prohibitivo, de manera que sería útil realizar una selección previa de los efectos finales aparentemente más sensibles. Sin embargo, en algunos casos no es fácil identificar los más sensibles mediante una inspección visual, de manera que hay que aplicar el modelo a todos ellos.

- Se debería notificar siempre la dosis de referencia y su límite inferior de confianza, de manera que la calidad de los datos y el ajuste del modelo sean claros y se puedan comparar sus potencias basándose en la dosis de referencia.

- Se deben presentar los resultados de los distintos métodos utilizados en los modelos de dosis-respuesta.

3. Recomendaciones

- Se deben examinar los protocolos de las pruebas de toxicidad (por ejemplo, las directrices de la Organización de Cooperación y Desarrollo Económicos) para conseguir unos resultados óptimos de las dosis de referencia y demás métodos basados en modelos de la relación dosis-respuesta, por ejemplo las formulaciones óptimas correspondientes al número de animales y el número de dosis para diferentes curvas de la relación dosis-respuesta.

- Hay que elaborar mejores orientaciones para el análisis combinado de distintas series de datos, a fin de estimar las dosis de referencia con mayor precisión.

- Es necesario fomentar un mayor conocimiento de cuándo y cómo se ha de utilizar la respuesta de referencia.

- Hay que tratar de conocer mejor la forma de la curva de la relación dosis-respuesta a dosis bajas. Se requieren nuevas investigaciones para determinar la base biológica de la extrapolación (por ejemplo, utilizando biomarcadores, precursores de tumores, animales modificados genéticamente y la tóxicocinética para la estimación de dosis específicas).

- Es necesario elaborar orientaciones mejores para la comunicación del riesgo basada en los resultados de los modelos dosis-respuesta y de las técnicas de evaluación probabilística. Deben incluir la comunicación de los tipos de incertidumbre y la relación con la variabilidad estadística, la imprecisión y la utilización de intervalos de confianza.

- Se debe examinar la utilización de modelos de la relación dosis-respuesta y se han de elaborar principios generales adicionales para su uso cuando se disponga de más experiencia.

THE ENVIRONMENTAL HEALTH CRITERIA SERIES (continued)

Acetaldehyde (No. 167, 1995)
Acetone (No. 207, 1998)
Acetonitrile (No. 154, 1993)
Acrolein (No. 127, 1991)
Acrylamide (No. 49, 1985)
Acrylic acid (No. 191, 1997)
Acrylonitrile (No. 28, 1983)
Aged population, principles for evaluating the effects of chemicals (No. 144, 1992)
Aldicarb (No. 121, 1991)
Aldrin and dieldrin (No. 91, 1989)
Allergic hypersensitization associated with exposure to chemicals, principles and methods for assessing (No. 212, 1999)
Allethrins (No. 87, 1989)
Aluminium (No. 194, 1997)
Amitrole (No. 158, 1994)
Ammonia (No. 54, 1986)
Anticoagulant rodenticides (No. 175, 1995)
Arsenic (No. 18, 1981, 1st edition)
Arsenic and arsenic compounds (No. 224, 2001, 2nd edition)
Asbestos and other natural mineral fibres (No. 53, 1986)
Assessment of risks to human health from exposure to chemicals, principles for the (No. 210, 1999)
Autoimmunity associated with exposure to chemicals, principles and methods for assessing (No. 236, 2006)
Bacillus thuringiensis (No. 217, 1999)
Barium (No. 107, 1990)
Benomyl (No. 148, 1993)
Benzene (No. 150, 1993)
Beryllium (No. 106, 1990)
Biomarkers and risk assessment: concepts and principles (No. 155, 1993)
Biomarkers in risk assessment: validity and validation (No. 222, 2001)
Biotoxins, aquatic (marine and freshwater) (No. 37, 1984)
Boron (No. 204, 1998)
Brominated diphenylethers (No. 162, 1994)
Butanols – four isomers (No. 65, 1987)
Cadmium (No. 134, 1992)
Cadmium – environmental aspects (No. 135, 1992)
Camphechlor (No. 45, 1984)
Carbamate pesticides: a general introduction (No. 64, 1986)
Carbaryl (No. 153, 1994)
Carbendazim (No. 149, 1993)
Carbon disulfide (No. 10, 1979)
Carbon monoxide (No. 13, 1979, 1st edition) (No. 213, 1999, 2nd edition)
Carbon tetrachloride (No. 208, 1999)
Carcinogens, summary report on the evaluation of short-term *in vitro* tests (No. 47, 1985)
Carcinogens, summary report on the evaluation of short-term *in vivo* tests (No. 109, 1990)
Chlordane (No. 34, 1984)
Chlordimeform (No. 199, 1997)
Chlordecone (No. 43, 1984)
Chlorendic acid and anhydride (No. 185, 1996)
Chlorinated paraffins (No. 181, 1996)
Chlorine and hydrogen chloride (No. 21, 1982)
Chloroalkyl ethers, selected (No. 201, 1998)
Chlorobenzenes other than hexachlorobenzene (No. 128, 1991)
Chlorofluorocarbons, fully halogenated (No. 113, 1990)
Chlorofluorocarbons, partially halogenated (ethane derivatives) (No. 139, 1992) (methane derivatives) (No. 126, 1991)
Chloroform (No. 163, 1994)
Chlorophenols (No. 93, 1989)
Chlorothalonil (No. 183, 1996)
Chromium (No. 61, 1988)
Chrysotile asbestos (No. 203, 1998)
Copper (No. 200, 1998)
Cresols (No. 168, 1995)
Cyhalothrin (No. 99, 1990)
Cypermethrin (No. 82, 1989)
Cypermethrin, alpha- (No. 142, 1992)
DDT and its derivatives (No. 9, 1979)
DDT and its derivatives – environmental aspects (No. 83, 1989)
Deltamethrin (No. 97, 1990)
Demeton-S-methyl (No. 197, 1997)
Dermal absorption (No. 235, 2006)
Diaminotoluenes (No. 74, 1987)
Diazinon (No. 198, 1997)
1,2-Dibromoethane (No. 177, 1996)
Di-*n*-butyl phthalate (No. 189, 1997)
1,2-Dichloroethane (No. 62, 1987, 1st edition) (No. 176, 1995, 2nd edition)
2,4-Dichlorophenoxyacetic acid (2,4-D) (No. 29, 1984)
2,4-Dichlorophenoxyacetic acid – environmental aspects (No. 84, 1989)
1,3-Dichloropropene, 1,2-dichloropropane and mixtures (No. 146, 1993)
Dichlorvos (No. 79, 1988)
Dinitro-ortho-cresol (No. 220, 2000)
Diesel fuel and exhaust emissions (No. 171, 1996)
Diethylhexyl phthalate (No. 131, 1992)
Diflubenzuron (No. 184, 1996)
Dimethoate (No. 90, 1989)
Dimethylformamide (No. 114, 1991)
Dimethyl sulfate (No. 48, 1985)
Diseases of suspected chemical etiology and their prevention, principles of studies on (No. 72, 1987)
Disinfectants and disinfectant by-products (No. 216, 1999)
Dithiocarbamate pesticides, ethylenethiourea, and propylenethiourea: a general introduction (No. 78, 1988)
Electromagnetic fields (No. 137, 1992)
Elemental speciation in human health risk assessment (No. 234, 2006)
Endosulfan (No. 40, 1984)
Endrin (No. 130, 1992)
Environmental epidemiology, guidelines on studies in (No. 27, 1983)
Epichlorohydrin (No. 33, 1984)
Essential trace elements: Principles and

THE ENVIRONMENTAL HEALTH CRITERIA SERIES (continued)

methods for the assessment of risk (No. 228, 2001)
Ethylbenzene (No. 186, 1996)
Ethylene oxide (No. 55, 1985)
Extremely low frequency (ELF) fields (No. 36, 1984)
(No. 238, 2007)
Fenitrothion (No. 133, 1992)
Fenvalerate (No. 95, 1990)
Flame retardants: a general introduction (No. 192, 1997)
Flame retardants: tris(chloropropyl) phosphate and tris(2-chloroethyl) phosphate (No. 209, 1998)
Flame retardants: tris(2-butoxyethyl) phosphate, tris(2-ethylhexyl) phosphate and tetrakis(hydroxymethyl) phosphonium salts (No. 218, 2000)
Fluorides (No. 227, 2001)
Fluorine and fluorides (No. 36, 1984)
Food additives and contaminants in food, principles for the safety assessment of (No. 70, 1987)
Formaldehyde (No. 89, 1989)
Fumonisin B_1 (No. 219, 2000)
Genetic effects in human populations, guidelines for the study of (No. 46, 1985)
Glyphosate (No. 159, 1994)
Guidance values for human exposure limits (No. 170, 1994)
Health risks in children associated with exposure to chemicals, principles for evaluating (No. 237, 2006)
Heptachlor (No. 38, 1984)
Hexachlorobenzene (No. 195, 1997)
Hexachlorobutadiene (No. 156, 1994)
Alpha- and beta-hexachlorocyclohexanes (No. 123, 1992)
Hexachlorocyclopentadiene (No. 120, 1991)
n-Hexane (No. 122, 1991)
Human exposure assessment (No. 214, 2000)
Hydrazine (No. 68, 1987)
Hydrogen sulfide (No. 19, 1981)
Hydroquinone (No. 157, 1994)
Immunotoxicity associated with exposure to chemicals, principles and methods for assessment (No. 180, 1996)
Infancy and early childhood, principles for evaluating health risks from chemicals during (No. 59, 1986)
Isobenzan (No. 129, 1991)
Isophorone (No. 174, 1995)
Kelevan (No. 66, 1986)
Lasers and optical radiation (No. 23, 1982)
Lead (No. 3, 1977)[a]
Lead, inorganic (No. 165, 1995)
Lead – environmental aspects (No. 85, 1989)
Lindane (No. 124, 1991)
Linear alkylbenzene sulfonates and related compounds (No. 169, 1996)
Magnetic fields (No. 69, 1987)
Man-made mineral fibres (No. 77, 1988)
Manganese (No. 17, 1981)
Mercury (No. 1, 1976)[a]
Mercury – environmental aspects (No. 86, 1989)
Mercury, inorganic (No. 118, 1991)
Methanol (No. 196, 1997)
Methomyl (No. 178, 1996)
2-Methoxyethanol, 2-ethoxyethanol, and their acetates (No. 115, 1990)
Methyl bromide (No. 166, 1995)
Methylene chloride (No. 32, 1984, 1st edition)
(No. 164, 1996, 2nd edition)
Methyl ethyl ketone (No. 143, 1992)
Methyl isobutyl ketone (No. 117, 1990)
Methylmercury (No. 101, 1990)
Methyl parathion (No. 145, 1992)
Methyl *tertiary*-butyl ether (No. 206, 1998)
Mirex (No. 44, 1984)
Modelling dose–response for the risk assessment of chemicals, Principles for (No. 239, 2009)
Morpholine (No. 179, 1996)
Mutagenic and carcinogenic chemicals, guide to short-term tests for detecting (No. 51, 1985)
Mycotoxins (No. 11, 1979)
Mycotoxins, selected: ochratoxins, trichothecenes, ergot (No. 105, 1990)
Nephrotoxicity associated with exposure to chemicals, principles and methods for the assessment of (No. 119, 1991)
Neurotoxicity associated with exposure to chemicals, principles and methods for the assessment of (No. 60, 1986)
Neurotoxicity risk assessment for human health, principles and approaches (No. 223, 2001)
Nickel (No. 108, 1991)
Nitrates, nitrites, and *N*-nitroso compounds (No. 5, 1978)[a]
Nitrobenzene (No. 230, 2003)
Nitrogen oxides (No. 4, 1977, 1st edition)[a]
(No. 188, 1997, 2nd edition)
2-Nitropropane (No. 138, 1992)
Nitro-and nitro-oxypolycyclic aromatic hydrocarbons, selected (No. 229, 2003)
Noise (No. 12, 1980)[a]
Organophosphorus insecticides: a general introduction (No. 63, 1986)
Palladium (No. 226, 2001)
Paraquat and diquat (No. 39, 1984)
Pentachlorophenol (No. 71, 1987)
Permethrin (No. 94, 1990)
Pesticide residues in food, principles for the toxicological assessment of (No. 104, 1990)
Petroleum products, selected (No. 20, 1982)
Phenol (No. 161, 1994)
d-Phenothrin (No. 96, 1990)
Phosgene (No. 193, 1997)
Phosphine and selected metal phosphides (No. 73, 1988)
Photochemical oxidants (No. 7, 1978)
Platinum (No. 125, 1991)
Polybrominated biphenyls (No. 152, 1994)
Polybrominated dibenzo-*p*-dioxins and dibenzofurans (No. 205, 1998)
Polychlorinated biphenyls and terphenyls (No. 2, 1976, 1st edition)[a]
(No. 140, 1992, 2nd edition)
Polychlorinated dibenzo-*p*-dioxins and dibenzofurans (No. 88, 1989)
Polycyclic aromatic hydrocarbons, selected non-heterocyclic (No. 202, 1998)
Progeny, principles for evaluating health risks associated with exposure to chemicals during pregnancy (No. 30, 1984)
1-Propanol (No. 102, 1990)
2-Propanol (No. 103, 1990)
Propachlor (No. 147, 1993)
Propylene oxide (No. 56, 1985)
Pyrrolizidine alkaloids (No. 80, 1988)
Quintozene (No. 41, 1984)
Quality management for chemical safety testing (No. 141, 1992)
Radiofrequency and microwaves (No. 16, 1981)
Radionuclides, selected (No. 25, 1983)
Reproduction, principles for evaluating health risks associated with exposure to chemicals (No. 225, 2001)
Resmethrins (No. 92, 1989)

---

[a] Out of print

THE ENVIRONMENTAL HEALTH CRITERIA SERIES (continued)

Selenium (No. 58, 1986)
Static fields (No. 232, 2006)
Synthetic organic fibres, selected (No. 151, 1993)
Styrene (No. 26, 1983)
Sulfur oxides and suspended particulate matter (No. 8, 1979)
Tecnazene (No. 42, 1984)
Tetrabromobisphenol A and derivatives (No. 172, 1995)
Tetrachloroethylene (No. 31, 1984)
Tetradifon (No. 67, 1986)
Tetramethrin (No. 98, 1990)
Thallium (No. 182, 1996)
Thiocarbamate pesticides: a general introduction (No. 76, 1988)
Tin and organotin compounds (No. 15, 1980)
Titanium (No. 24, 1982)
Tobacco use and exposure to other agents (No. 211, 1999)
Toluene (No. 52, 1986)
Toluene diisocyanates (No. 75, 1987)
Toxicity of chemicals (Part 1), principles and methods for evaluating the (No. 6, 1978)
Toxicokinetic studies, principles of (No. 57, 1986)
Transgenic animal mutagenicity assays (No. 233, 2006)
Tributyl phosphate (No. 112, 1991)
Tributyltin compounds (No. 116, 1990)
Trichlorfon (No. 132, 1992)
1,1,1-Trichloroethane (No. 136, 1992)
Trichloroethylene (No. 50, 1985)
Tricresyl phosphate (No. 110, 1990)
Triphenyl phosphate (No. 111, 1991)
Tris- and bis(2,3-dibromopropyl) phosphate (No. 173, 1995)
Ultrasound (No. 22, 1982)
Ultraviolet radiation (No. 14, 1979, 1st edition) (No. 160, 1994, 2nd edition)
Vanadium (No. 81, 1988)
Vinyl chloride (No. 215, 1999)
Vinylidene chloride (No. 100, 1990)
White spirit (No. 187, 1996)
Xylenes (No. 190, 1997)
Zinc (No. 221, 2001)

CONCISE INTERNATIONAL CHEMICAL ASSESSMENT DOCUMENT SERIES

CICADs are IPCS risk assessment documents that provide concise but critical summaries of the relevant scientific information concerning the potential effects of chemicals upon human health and/or the environment

Acrolein (No. 43, 2002)
Acrylonitrile (No. 39, 2002)
Arsine: human health aspects (No. 47, 2002)
Asphalt (bitumen) (No. 59, 2004)
Azodicarbonamide (No. 16, 1999)
Barium and barium compounds (No. 33, 2001)
Benzoic acid and sodium benzoate (No. 26, 2000)
Benzyl butyl phthalate (No. 17, 1999)
Beryllium and beryllium compounds (No. 32, 2001)
Biphenyl (No. 6, 1999)
Bromoethane (No. 42, 2002)
1,3-Butadiene: human health aspects (No. 30, 2001)
2-Butenal (No. 74, 2008)
2-Butoxyethanol (No. 10, 1998)
2-Butoxyethanol (update) (No. 67, 2005)
Butyl acetates (No. 64, 2005)
Carbon disulfide (No. 46, 2002)
Chloral hydrate (No. 25, 2000)
Chlorinated naphthalenes (No. 34, 2001)
Chlorine dioxide (No. 37, 2002)
4-Chloroaniline (No. 48, 2003)
Chlorobenzenes other than hexachlorobenzene: environmental aspects (No. 60, 2004)
Chloroform (No. 58, 2004)
Coal tar creosote (No. 62, 2004)
Cobalt and inorganic cobalt compounds (No. 69, 2006)
Crystalline silica, quartz (No. 24, 2000)
Cumene (No. 18, 1999)
1,2-Diaminoethane (No. 15, 1999)
3,3'-Dichlorobenzidine (No. 2, 1998)
1,2-Dichloroethane (No. 1, 1998)
1,1-Dichloroethene (Vinylidene chloride) (No. 51, 2003)
2,2-Dichloro-1,1,1-trifluoroethane (HCFC-123) (No. 23, 2000)
Diethyl phthalate (No. 52, 2003)
Diethylene glycol dimethyl ether (No. 41, 2002)
Dimethylformamide, N,N- (No. 31, 2001)
Diphenylmethane diisocyanate (MDI) (No. 27, 2001)
Elemental mercury and inorganic mercury compounds; human health aspects (No. 50, 2004)
Ethylenediamine (No. 15, 1999)
Ethylene glycol: environmental aspects (No. 22, 2000)
Ethylene glycol: human health aspects (No. 45, 2002))
Ethylene oxide (No. 54, 2003)
Formaldehyde (No. 40, 2002)
2-Furaldehyde (No. 21, 2000)
Glyoxal (No. 57, 2004)
Heptachlor (No. 70, 2006)

CONCISE INTERNATIONAL CHEMICAL ASSESSMENT DOCUMENT SERIES (continued)

HCFC-123 (No. 23, 2000)
Hydrogen cyanide and cyanides: human health aspects (No. 61, 2004)
Hydrogen sulfide: human health aspects (No. 53, 2003)
Limonene (No. 5, 1998)
Manganese and its compounds (No. 12, 1999)
Manganese and its compounds: environmental aspects (No. 63, 2004)
Mercury, elemental, and inorganic mercury compounds: human health aspects (No. 50, 2003)
Methyl and ethyl cyanoacrylates (No. 36, 2001)
Methyl chloride (No. 28, 2001)
Methyl methacrylate (No. 4, 1998)
N-Methyl-2-pyrrolidone (No. 35, 2001)
Methyltin, butyltin, and octyltin compounds, Mono- and disubstituted (No. 73, 2006)
Mononitrophenols (No. 20, 2000)
N-Nitrosodimethylamine (No. 38, 2002)
Phenylhydrazine (No. 19, 2000)
N-Phenyl-1-naphthylamine (No. 9, 1998)
Polychlorinated biphenyls: human health aspects (No. 55, 2003)
Resorcinol (No. 71, 2006)
Silver and silver compounds: environmental aspects (No. 44, 2002)
1,1,2,2-Tetrachloroethane (No. 3, 1998)
Tetrachloroethene (No. 68, 2006)
1,1,2,2-Tetrafluoroethane (No. 11, 1998)
Thiourea (No. 49, 2003)
Tin and inorganic tin compounds (No. 65, 2005)
o-Toluidine (No. 7, 1998)
2,4,6-Tribromophenol and other simple brominated phenols (No. 66, 2005)
Tributylin oxide (No. 14, 1999)
1,2,3-Trichloropropane (No. 56, 2003)
Triglycidyl isocyanurate (No. 8, 1998)
Triphenyltin compounds (No. 13, 1999)
Vanadium pentoxide and other inorganic vanadium compounds (No. 29, 2001)

To order further copies of monographs in these series, please contact WHO Press, World Health Organization, 1211 Geneva 27, Switzerland (Fax: +41-22-791 4857; E-mail: bookorders@who.int)

ISBN 978 92 4 157239 2