

# IPCS

INTERNATIONAL PROGRAMME ON CHEMICAL SAFETY

**IPCS Harmonization Project**

## **Guidance Document on Evaluating and Expressing Uncertainty in Hazard Characterization**

SECOND EDITION 2017

### **IOMC**

INTER-ORGANIZATION PROGRAMME FOR THE SOUND MANAGEMENT OF CHEMICALS  
A cooperative agreement among FAO, ILO, UNEP, UNIDO, UNITAR, WHO, World Bank 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 World Health Organization, the International Labour Organization or the United Nations Environment Programme.

## **Harmonization Project Document 11**

### **GUIDANCE DOCUMENT ON EVALUATING AND EXPRESSING UNCERTAINTY IN HAZARD CHARACTERIZATION**

This project was conducted within the IPCS project on the Harmonization of Approaches to the Assessment of Risk from Exposure to Chemicals.

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

SECOND EDITION 2017



The **International Programme on Chemical Safety (IPCS)** was established in 1980. 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.

This publication was developed in the IOMC context. The contents do not necessarily reflect the views or stated policies of individual IOMC Participating Organizations.

The **Inter-Organization Programme for the Sound Management of Chemicals (IOMC)** was established in 1995 following recommendations made by the 1992 UN Conference on Environment and Development to strengthen cooperation and increase international coordination in the field of chemical safety. The Participating Organizations are: FAO, ILO, UNDP, UNEP, UNIDO, UNITAR, WHO, World Bank and OECD. 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.

Guidance document on evaluating and expressing uncertainty in hazard characterization – 2<sup>nd</sup> edition

(IPCS harmonization project document ; no. 11)

ISBN 978-92-4-151354-8

© **World Health Organization, 2018**

Some rights reserved. This work is available under the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 IGO licence (CC BY-NC-SA 3.0 IGO; <https://creativecommons.org/licenses/by-nc-sa/3.0/igo>).

Under the terms of this licence, you may copy, redistribute and adapt the work for non-commercial purposes, provided the work is appropriately cited, as indicated below. In any use of this work, there should be no suggestion that WHO endorses any specific organization, products or services. The use of the WHO logo is not permitted. If you adapt the work, then you must license your work under the same or equivalent Creative Commons licence. If you create a translation of this work, you should add the following disclaimer along with the suggested citation: “This translation was not created by the World Health Organization (WHO). WHO is not responsible for the content or accuracy of this translation. The original English edition shall be the binding and authentic edition”.

Any mediation relating to disputes arising under the licence shall be conducted in accordance with the mediation rules of the World Intellectual Property Organization.

**Suggested citation.** Guidance document on evaluating and expressing uncertainty in hazard characterization– 2<sup>nd</sup> edition. Geneva: World Health Organization; 2018. Licence: [CC BY-NC-SA 3.0 IGO](https://creativecommons.org/licenses/by-nc-sa/3.0/igo).

**Cataloguing-in-Publication (CIP) data.** CIP data are available at <http://apps.who.int/iris>.

**Sales, rights and licensing.** To purchase WHO publications, see <http://apps.who.int/bookorders>. To submit requests for commercial use and queries on rights and licensing, see <http://www.who.int/about/licensing>.

**Third-party materials.** If you wish to reuse material from this work that is attributed to a third party, such as tables, figures or images, it is your responsibility to determine whether permission is needed for that reuse and to obtain permission from the copyright holder. The risk of claims resulting from infringement of any third-party-owned component in the work rests solely with the user.

**General disclaimers.** The designations employed and the presentation of the material in this publication do not imply the expression of any opinion whatsoever on the part of WHO concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. Dotted and dashed lines on maps represent approximate border lines for which there may not yet be full agreement.

The mention of specific companies or of certain manufacturers' products does not imply that they are endorsed or recommended by WHO in preference to others of a similar nature that are not mentioned. Errors and omissions excepted, the names of proprietary products are distinguished by initial capital letters.

All reasonable precautions have been taken by WHO to verify the information contained in this publication. However, the published material is being distributed without warranty of any kind, either expressed or implied. The responsibility for the interpretation and use of the material lies with the reader. In no event shall WHO be liable for damages arising from its use.

This publication contains the collective views of an international group of experts and does not necessarily represent the decisions or the policies of WHO.

Printed in Switzerland

## TABLE OF CONTENTS

FOREWORD .....	viii
FOREWORD TO THE SECOND EDITION 2018.....	ix
ACKNOWLEDGEMENTS.....	x
CONTRIBUTORS.....	xi
LIST OF ABBREVIATIONS.....	xvii
EXECUTIVE SUMMARY .....	xix
<b>1. INTRODUCTION .....</b>	<b>1</b>
1.1 Purpose and scope.....	1
1.2 Use of quantitative evaluation of uncertainties in hazard characterization to achieve better informed and more transparent decisions.....	2
1.3 Organization of this document.....	4
<b>2. CONTEXT AND OVERVIEW .....</b>	<b>6</b>
2.1 Uncertainty in hazard characterization evaluated in the context of problem formulation for risk assessment .....	6
2.2 Evaluation of uncertainty in hazard characterization in concert with other risk assessment methodologies .....	7
2.3 Overview of the principles and framework for evaluating uncertainty in hazard characterization .....	9
2.3.1 Principle 1: Individual-level effects (magnitude) and population-level effects (incidence) are conceptually distinct.....	9
2.3.2 Principle 2: For all types of end-points, the magnitude of effect $M$ can be regarded as changing gradually .....	9
2.3.3 Principle 3: The concept of an “effect metric” for $M$ forms the basis of “equipotency” and differences in “sensitivity” .....	10
2.3.4 Principle 4: Making inferences from a point of departure involves making adjustments and accounting for variability and uncertainty .....	10
2.3.5 Applying the framework.....	11
2.4 Implications for problem formulation and interaction with risk management .....	12
2.4.1 The target human dose $HD_M^I$ as the focus of hazard characterization when evaluating uncertainties .....	12
2.4.1.1 A conceptual transition to the $HD_M^I$ .....	12
2.4.1.2 Transition from the deterministic RfD to the $HD_M^I$ as a natural extension of the transition from the NOAEL to the BMD.....	12
2.4.2 Being “conservative” and its relationship to evaluating uncertainty .....	15
2.4.2.1 “Conservative” in relation to protection goals.....	15
2.4.2.2 “Conservative” in relation to uncertainty: “coverage” and “degree of uncertainty” .....	15
2.4.3 Using conclusions from analysis of uncertainties to inform the question, “Do we know enough?” .....	17
2.4.4 How to adapt the approach for different regulatory contexts .....	18

<b>3.</b>	<b>DETAILED DESCRIPTION OF THE APPROACH TO EVALUATING UNCERTAINTY IN HAZARD CHARACTERIZATION.....</b>	<b>19</b>
3.1	Aspects of hazard characterization and associated uncertainties.....	19
3.2	A framework for evaluating uncertainty in hazard characterization based on four fundamental principles.....	21
3.2.1	Principle 1: Individual-level effects (magnitude) and population-level effects (incidence) are conceptually distinct.....	21
3.2.2	Principle 2: For all types of end-points, the magnitude of effect <i>M</i> can be regarded as changing gradually .....	22
3.2.2.1	Continuous end-points .....	22
3.2.2.2	Quantal end-points .....	23
3.2.3	Principle 3: The concept of an “effect metric” for <i>M</i> forms the basis of “equipotency” and differences in “sensitivity” .....	26
3.2.4	Principle 4: Making inferences from a point of departure involves making adjustments while accounting for uncertainty and variability .....	27
3.2.4.1	Uncertainty versus variability .....	27
3.2.4.2	Adjustment and uncertainty .....	28
3.3	Applying the framework .....	28
3.3.1	Quantifying uncertainties individually and in combination.....	28
3.3.1.1	Quantifying uncertainties individually .....	28
3.3.1.2	Combining uncertainties .....	29
3.3.2	Spreadsheet tool “APROBA” for facilitating approximate probabilistic analysis.....	30
3.4	Evaluation of other uncertainties .....	33
3.4.1	Uncertainties regarding distributions used in the probabilistic approach.....	33
3.4.2	Uncertainties regarding the assumption of independence .....	33
3.4.3	Uncertainties relating to qualitative aspects of hazard assessment.....	34
3.4.4	Uncertainties that are difficult to quantify .....	34
<b>4.</b>	<b>IMPLEMENTATION OF THE APPROACH: PRELIMINARY UNCERTAINTY DISTRIBUTIONS .....</b>	<b>36</b>
4.1	Using historical data: general approach.....	36
4.2	Points of departure .....	37
4.2.1	Benchmark dose.....	37
4.2.2	No-observed-adverse-effect level .....	37
4.2.3	Lowest-observed-adverse-effect level .....	39
4.3	Exposure duration .....	39
4.4	Interspecies extrapolation .....	40
4.4.1	Body size adjustment .....	41
4.4.1.1	Adjustment of oral doses .....	41
4.4.1.2	Uncertainty distribution for allometric scaling factor for oral doses.....	41
4.4.1.3	Body size adjustment of inhalation exposures and its uncertainty .....	41
4.4.2	Toxicokinetic/toxicodynamic differences.....	42
4.5	Human interindividual differences in sensitivity.....	43
4.5.1	Background on uncertainties in human interindividual differences .....	43
4.5.2	Uncertainty distributions for intraspecies variability.....	43
4.6	Route-to-route extrapolation.....	46

4.7	Summary of generic uncertainties per aspect .....	47
4.8	Evaluating primary uncertainty for other aspects .....	47
4.9	Chemical-specific distributions for primary uncertainties.....	49
4.10	Evaluating secondary uncertainties .....	50
<b>5.</b>	<b>ILLUSTRATION OF THE APPROACH USING GENERIC HAZARD CHARACTERIZATION SCENARIOS .....</b>	<b>54</b>
5.1	Probabilistic uncertainty characterization of the deterministic reference dose and derivation of a probabilistic reference dose .....	54
5.2	Sensitivity analysis: ranking the sources of uncertainty .....	57
5.3	Visualization of the uncertainty in the target human dose for different values of coverage and population incidence .....	58
<b>6.</b>	<b>INTERPRETATION AND USE OF RESULTS.....</b>	<b>60</b>
6.1	Setting a health-based guidance value .....	61
6.2	Acceptability and communication of health risks for a given situation.....	64
6.3	Prioritization of chemicals based on predicted risks.....	66
6.4	Socioeconomic analysis .....	66
<b>7.</b>	<b>NEXT STEPS .....</b>	<b>68</b>
7.1	Further research .....	68
	7.1.1 Methodological issues.....	68
	7.1.2 Input data .....	69
7.2	Implementation and dissemination .....	70
<b>8.</b>	<b>REFERENCES.....</b>	<b>72</b>
<b>9.</b>	<b>GLOSSARY OF TERMS.....</b>	<b>76</b>
	<b>ANNEX 1: COMBINING UNCERTAINTIES FOR THE THREE APPROACHES....</b>	<b>81</b>
A1.1	General approach to combining uncertainties.....	81
A1.2	Non-probabilistic approach to combining uncertainties .....	82
A1.3	Approximate probabilistic approach to combining uncertainties .....	83
A1.4	Full probabilistic approach to combining uncertainties.....	85
	<b>ANNEX 2: APROBA SPREADSHEET TOOL USER GUIDE .....</b>	<b>87</b>
A2.1	Introduction.....	87
A2.2	General layout of worksheets.....	87
A2.3	Step-by-step procedure for using APROBA.....	87
A2.4	Using intermediate calculations to estimate each aspect’s contributions to uncertainty.....	91
A2.5	Visualizing the impact of changing coverage and incidence <i>I</i> .....	91

**ANNEX 3: DETAILS OF ESTIMATING UNCERTAINTY DISTRIBUTIONS FOR USE OF A NO-OBSERVED-ADVERSE-EFFECT LEVEL, EXPOSURE DURATION, INTERSPECIES EXTRAPOLATION AND ROUTE-TO-ROUTE EXTRAPOLATION .....93**

A3.1 Point of departure: uncertainty in using a NOAEL as a surrogate for a BMD .....93  
     A3.1.1 Continuous end-points .....93  
     A3.1.2 Quantal end-points .....94  
 A3.2 Exposure duration: uncertainty in using a subchronic or subacute study as a surrogate for a chronic study.....95  
     A3.2.1 Subchronic to chronic extrapolation .....95  
     A3.2.2 Subacute to chronic extrapolation.....97  
 A3.3 Interspecies extrapolation: uncertainty in using a test animal species as a surrogate for humans (after body size adjustment).....98  
 A3.4 Route-to-route extrapolation: uncertainty in using an oral study as a basis for inhalation hazard, or vice versa .....100  
 A3.5 References .....101

**ANNEX 4: DETAILED RESULTS OF ANALYSES OF HISTORICAL DATA ON INTRASPECIES HUMAN VARIABILITY .....104**

A4.1 Introduction.....104  
 A4.2 Observed intraspecies variation in toxicokinetics based on historical data.....105  
 A4.3 Observed intraspecies variation in toxicodynamics based on historical data .....107  
 A4.4 Estimated uncertainty in overall intraspecies variability .....113  
 A4.5 References .....115

**ANNEX 5: CASE-STUDY: DEOXYNIVALENOL (DON).....116**

A5.1 Introduction.....116  
     A5.1.1 General note .....116  
     A5.1.2 Problem formulation .....116  
     A5.1.3 General approach .....116  
 A5.2 Tier 1: Hazard characterization for general toxicity after repeated dosing using the NOAEL and default assessment factors.....118  
     A5.2.1 Hazard characterization aspects .....118  
         A5.2.1.1 Aspect I: Setting the PoD.....118  
         A5.2.1.2 Aspect II: Adjustment of the PoD (LOAEL à NOAEL) .....118  
         A5.2.1.3 Aspect III: Interspecies extrapolation .....119  
         A5.2.1.4 Aspect IV: Intraspecies extrapolation.....119  
         A5.2.1.5 Assessment factors and derivation of the reference dose .....119  
     A5.2.2 Evaluation of uncertainties .....119  
         A5.2.2.1 Inputs related to study, end-point and protection goal.....120  
         A5.2.2.2 Inputs related to adjustment, variability and uncertainty.....121  
         A5.2.2.3 Results.....122  
         A5.2.2.4 Is a higher-tier assessment needed? .....123  
         A5.2.2.5 How to proceed? .....125  
 A5.3 Tier 2: Hazard characterization for general toxicity after repeated dosing based on the BMDL .....126  
     A5.3.1 Dose–response analysis for general toxicity (BMD approach) .....126  
     A5.3.2 BMDL for general toxicity .....126

A5.3.3	Evaluation of uncertainties .....	127
A5.3.4	Is a higher-tier assessment needed? .....	129
A5.4	Including further end-points .....	131
A5.4.1	Tier 1 assessment of end-points of reproductive toxicity using the NOAEL approach and default assessment factors.....	132
A5.4.1.1	Candidate studies .....	132
A5.4.1.2	Assessment factors.....	133
A5.4.1.3	Derivation of tier 1 reference doses .....	134
A5.4.1.4	Evaluation of uncertainties .....	134
A5.4.1.5	Is a higher-tier assessment needed? .....	136
A5.4.2	Tier 2 (BMD-based) evaluation of uncertainties .....	138
A5.4.2.1	Prenatal development.....	138
A5.4.2.2	Fertility.....	139
A5.4.2.3	Assessment factors and derivation of the reference dose .....	139
A5.4.2.4	Evaluation of uncertainties .....	140
A5.4.2.5	Is a higher-tier assessment needed? .....	141
A5.5	Conclusions and further options .....	143
A5.6	References.....	143
Appendix I:	Dose–response information considered in the case-study .....	145
Appendix II:	Results of BMD analyses .....	147

## FOREWORD

For ethical reasons, the potential health effects associated with exposure of humans to toxic substances cannot normally be assessed directly in a planned experimental setting. Instead, the process of evaluating human health effects usually needs to rely on data that are only indirectly relevant. For instance, the data may relate to another species, to an exposure scenario that differs from the one applicable to humans or to a specific subpopulation of humans with many interindividual differences, apart from exposure to the chemical of interest. Therefore, the process of evaluating human health effects as a function of (potential) exposure, denoted as hazard characterization, necessarily involves uncertainties associated with extrapolating indirectly relevant results to humans. Ignoring these uncertainties may lead to incomplete risk assessments as well as suboptimal decision-making and risk communication. Risk assessors have to take uncertainty explicitly into account. Effective risk management does not require the elimination of uncertainty; rather, it requires that any such uncertainty has been made visible and taken into consideration.<sup>1</sup>

Uncertainty and variability in exposure assessment of chemicals have been addressed in the International Programme on Chemical Safety's (IPCS) *Guidance document on characterizing and communicating uncertainty in exposure assessment*, published in 2008.<sup>2</sup> In 2007, the Harmonization Project Steering Committee decided to launch a project group with the aim of developing guidance on the evaluation and expression of uncertainty in hazard characterization as well, resulting in the present document.

This document is part of a series of coordinated international, regional and national efforts on the assessment and management of hazardous chemicals, the main impetus for which arose from the 1992 United Nations Conference on Environment and Development (UNCED). The commitment by governments was reconfirmed at the 2002 World Summit on Sustainable Development and in 2006 in the Strategic Approach to International Chemicals Management (SAICM). The IPCS project on the Harmonization of Approaches to the Assessment of Risk from Exposure to Chemicals (Harmonization Project) is conducted under UNCED Agenda 21, Chapter 19, and contributes to the implementation of SAICM. In particular, the project addresses the SAICM objective on Risk Reduction and the SAICM Global Plan of Action activity to “develop and use new and harmonized methods for risk assessment”.

The IPCS Harmonization Project goal is *to improve chemical risk assessment globally, through the pursuit of common principles and approaches, and, hence, strengthen national and international management practices that deliver better protection of human health and the environment within the framework of sustainability*. The Harmonization Project aims to harmonize global approaches to chemical risk assessment, including by developing international guidance documents on specific issues. The guidance is intended for adoption and use in countries and by international bodies in the performance of chemical risk assessments. The guidance is developed by engaging experts worldwide. The project has been implemented using a stepwise approach, first sharing information and increasing understanding of methods and practices used by various countries, identifying areas where

---

<sup>1</sup> Funtowicz SO, Ravetz JR (1990). *Uncertainty and quality in science for policy*. Dordrecht: Kluwer Academic Publishers.

<sup>2</sup> IPCS (2008). *Guidance document on characterizing and communicating uncertainty in exposure assessment*. In: *Uncertainty and data quality in exposure assessment*. Geneva: World Health Organization, International Programme on Chemical Safety (Harmonization Project Document No. 6).

convergence of different approaches would be beneficial and then developing guidance that enables implementation of harmonized approaches. The project uses a building block approach, focusing at any one time on the aspects of risk assessment that are particularly important for harmonization. The project enables risk assessments (or components thereof) to be performed using internationally accepted methods, and these assessments can then be shared to avoid duplication and optimize use of valuable resources for risk management. It also promotes sound science as a basis for risk management decisions, promotes transparency in risk assessment and reduces unnecessary testing of chemicals. Advances in scientific knowledge can be translated into new harmonized methods.

This document is bold in laying out a vision and a road map, but incremental in building upon current typical assessment practices. This guidance largely focuses on how to quantitatively evaluate uncertainties in current assessment practices. As experience is gained with probabilistic approaches of hazard characterization, appropriate changes to routine practices may well evolve over time. It is emphasized that the probabilistic approach uses the same basic underlying scientific data that are used in any other type of hazard characterization, and in that sense it does not reduce uncertainties in risk estimates, per se. What it does is to quantitatively make visible uncertainties associated with the outcome of any given hazard characterization. This information may lead to better informed risk management decisions. Reducing uncertainty in risk estimates is achieved by generating additional information (data or more advanced methods). The probabilistic approach is, however, helpful in deciding what type of additional information would be potentially effective in reducing the uncertainty in the outcome from an upgraded hazard characterization.

The guidance offered here is consistent with the IPCS Harmonization Project monograph on *IPCS risk assessment terminology*.<sup>3</sup>

## FOREWORD TO THE SECOND EDITION 2017

After publication of the first version of this document a calculation error in the tier 1 assesment step of the DON case-study (ANNEX V) was noted. As a result of that error, calculation results based on the LOAEL indicated a greater discrepancy between the traditional “deterministic” and the (approximate) probabilistic assessment than is actually the case. While the general conclusions from tier 1 or the case-study as a whole are not affected. WHO nevertheless apologizes for any inconvenience caused by this error which has now been corrected in the new version. In this context WHO also took the opportunity to carry out some minor editorial revisions both in this document and in the APROBA spreadsheet tool. In addition, after publication of the first version of this guidance document the new approach was also the subject of a publication in the peer-reviewed literature (Chiu & Slob, 2015) which has now been added to the reference list (section 8).

---

<sup>3</sup> IPCS (2004). *IPCS risk assessment terminology*. Geneva: World Health Organization, International Programme on Chemical Safety (Harmonization Project Document No. 1).

## ACKNOWLEDGEMENTS

This guidance document was prepared under the auspices of the IPCS project on the Harmonization of Approaches to the Assessment of Risk from Exposure to Chemicals through two expert meetings and a peer review workshop. The World Health Organization (WHO) gratefully acknowledges the contributions of the experts involved, the co-chairs of the meetings and teleconferences, D. Bussard and T. Vermeire, K. Hughes for her work in the Secretariat while on loan to WHO, the National Institute for Public Health and the Environment (RIVM) of the Netherlands for hosting two of the meetings, as well as the financial contribution of the United States Environmental Protection Agency (USEPA) towards the development of the guidance document.

1<sup>st</sup> edition 2014 technically and linguistically edited by Marla Sheffer, Ottawa, Canada

.

## CONTRIBUTORS

A draft working paper was prepared by experts from the National Institute for Public Health and the Environment (Rijksinstituut voor Volksgezondheid en Milieu, or RIVM) of the Netherlands and the United States Environmental Protection Agency (USEPA). The draft working paper was discussed at an expert meeting held at RIVM on 16–17 September 2010. Participants at this meeting were (\* denotes author of the draft working paper):

David Bussard\*, United States Environmental Protection Agency, Washington, DC, United States of America (USA) (*Co-Chair*)

Weihsueh Chiu\*, United States Environmental Protection Agency, Washington, DC, USA

Kenny S. Crump, Louisiana Tech University, Ruston, LA, USA

Andy Hart, Risk and Numerical Sciences Team, The Food and Environment Research Agency, Sand Hutton, York, England, United Kingdom

Dale Hattis, Clark University, Marsh Institute, Worcester, MA, USA

Matthias Herzler, Federal Institute for Risk Assessment (BfR), Berlin, Germany

Daniel Krewski, McLaughlin Centre for Population Health Risk Assessment, University of Ottawa, Ottawa, Ontario, Canada

Kevin McNally, Statistical Modelling Team, Mathematical Sciences Unit, Health and Safety Laboratory, Harpur Hill, Buxton, Derbyshire, England, United Kingdom

Salomon Sand, National Food Administration, Uppsala, Sweden

R. Woodrow Setzer, National Center for Computational Toxicology, United States Environmental Protection Agency, Research Triangle Park, NC, USA

Wout Slob\*, National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands

Theo Vermeire\*, National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands (*Co-Chair*)

Paul White\*, United States Environmental Protection Agency, Washington, DC, USA

Lauren Zeise, Office of Environmental Health Hazard Assessment, California Environmental Protection Agency, Oakland, CA, USA

### **Representatives**

Laurent Bodin, Toxicology Unit, French Agency for Food, Environmental and Occupational Health & Safety (Agence nationale de sécurité sanitaire [ANSES]), Maisons-Alfort, France

Tomas Öberg, Scientific Committee and Advisory Forum Unit, European Food Safety Authority, Parma, Italy

**Secretariat**

Carolyn Vickers, Chemical Safety, World Health Organization, Geneva, Switzerland

A draft guidance document was prepared from the draft working paper through a series of teleconferences involving the following experts and representatives, who met to further develop the document at an expert meeting held at the Pan American Health Organization in Washington, DC, USA, on 29–30 November 2011 and who subsequently prepared the peer review draft of the guidance document through a series of teleconferences:

David Bussard, United States Environmental Protection Agency, Washington, DC, USA  
(*Co-Chair*)

Weihsueh Chiu, United States Environmental Protection Agency, Washington, DC, USA

Vicki Dellarco,<sup>4</sup> Office of Pesticide Programs, United States Environmental Protection Agency, Washington, DC, USA

Andy Hart, Leader, Risk and Numerical Sciences Team, The Food and Environment Research Agency, Sand Hutton, York, England, United Kingdom

Matthias Herzler, Federal Institute for Risk Assessment (BfR), Berlin, Germany

Wout Slob, National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands

Theo Vermeire, National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands (*Co-Chair*)

Paul White, United States Environmental Protection Agency, Washington, DC, USA

Lauren Zeise, Office of Environmental Health Hazard Assessment, California Environmental Protection Agency, Oakland, CA, USA

**Representatives**

Bernard Bottex, Unit of the Scientific Committee / SC Secretariat, European Food Safety Authority, Parma, Italy

George Fotakis, European Chemicals Agency, Helsinki, Finland

**Secretariat**

Kathy Hughes, Health Canada, Ottawa, Ontario, Canada

Agnes Soares, WHO Regional Office for the Americas, Washington, DC, USA

---

<sup>4</sup> Participated in the Washington meeting only.

Carolyn Vickers, Chemical Safety, World Health Organization, Geneva, Switzerland

The revised draft guidance document underwent peer review at a WHO workshop held at RIVM, Bilthoven, the Netherlands, on 19–20 November 2013. Participants at the workshop were:

Sandra Baird, Massachusetts Department of Environmental Protection, Boston, MA, USA

Tara Barton-Maclaren, Hazard Methodology Division, Existing Substances Risk Assessment Bureau, Health Canada, Ottawa, Ontario, Canada

Alan Boobis, Centre for Pharmacology & Therapeutics, Department of Medicine, Imperial College London, London, England, United Kingdom

Bernard Bottex, Unit of the Scientific Committee / SC Secretariat, European Food Safety Authority, Parma, Italy

Guillaume Boulanger, Risk Assessment Department, French Agency for Food, Environmental and Occupational Health & Safety (ANSES), Maisons-Alfort, France

David Bussard, United States Environmental Protection Agency, Washington, DC, USA  
(*Co-Chair*)

Weihsueh Chiu, United States Environmental Protection Agency, Washington, DC, USA

Peter Craig, Department of Mathematical Sciences, University of Durham, Durham, England, United Kingdom

Jean-Lou Dorne, Unit of the Scientific Committee / SC Secretariat, European Food Safety Authority, Parma, Italy

George Fotakis, European Chemicals Agency, Helsinki, Finland

Sandrine Fraize-Frontier, Risk Assessment Department, French Agency for Food, Environmental and Occupational Health & Safety (ANSES), Maisons-Alfort, France

Helen Goeden, Health Risk Assessment Unit, Minnesota Department of Health, St Paul, MN, USA

Matthias Greiner, Scientific Services, Federal Institute for Risk Assessment (BfR), Berlin, Germany

Anthony Hardy, Copmanthorpe, York, England, United Kingdom

Andy Hart, Risk and Numerical Sciences Team, The Food and Environment Research Agency, Sand Hutton, York, England, United Kingdom

Dale Hattis, Clark University, Marsh Institute, Worcester, MA, USA

Arie Havelaar,<sup>5</sup> National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands

Matthias Herzler, Federal Institute for Risk Assessment (BfR), Berlin, Germany

Johannes Linders, Hoogland, the Netherlands

Bette Meek, Chemical Risk Assessment, McLaughlin Centre for Population Health Risk Assessment, University of Ottawa, Ottawa, Ontario, Canada

Angelo Moretto, Department of Biomedical and Clinical Sciences, University of Milan, International Centre for Pesticides and Health Risk Prevention, Milan, Italy

Birgit Noerrung, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

Kerry Nugent, Existing Chemicals Program, National Industrial Chemicals Notification & Assessment Scheme, Australian Government Department of Health & Ageing, Surry Hills, NSW, Australia

Tomas Öberg, European Chemicals Agency, Helsinki, Finland

Christophe Rousselle, Chemical Risk Assessment Unit, French Agency for Food, Occupational and Environmental Health & Safety (ANSES), Maisons-Alfort, France

Wout Slob, National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands

Jacob van Klaveren,<sup>6</sup> National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands

Theo Vermeire, National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands (*Co-Chair*)

Christopher Weis, National Institute for Environmental Health Science, National Institutes of Health, United States Department of Health and Human Services, Bethesda, MD, USA

### **Nongovernmental organizations in official relations with WHO**

Alan Poole, European Centre for Ecotoxicology and Toxicology of Chemicals (ECETOC), Brussels, Belgium

### **Secretariat**

Kathy Hughes, Health Canada, Ottawa, Ontario, Canada

Carolyn Vickers, Chemical Safety, World Health Organization, Geneva, Switzerland

---

<sup>5</sup> Participated on first day only.

<sup>6</sup> Participated on second day only.

Following the peer review workshop, an expert meeting took place on 21–22 November 2013 to consider the input provided and to agree on the amendments to be made to finalize the guidance document. Participants at this final expert meeting were:

David Bussard, United States Environmental Protection Agency, Washington, DC, USA  
(*Co-Chair*)

Weihsueh Chiu, United States Environmental Protection Agency, Washington, DC, USA

Andy Hart, Risk and Numerical Sciences Team, The Food and Environment Research Agency, Sand Hutton, York, England, United Kingdom

Dale Hattis, Clark University, Marsh Institute, Worcester, MA, USA

Matthias Herzler, Federal Institute for Risk Assessment (BfR), Berlin, Germany

Wout Slob, National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands

Theo Vermeire, National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands (*Co-Chair*)

### **Representatives**

George Fotakis, European Chemicals Agency, Helsinki, Finland

### **Secretariat**

Kathy Hughes, Health Canada, Ottawa, Ontario, Canada

Carolyn Vickers, Chemical Safety, World Health Organization, Geneva, Switzerland

The document was subsequently finalized by a group of experts, with input from representatives of the European Chemicals Agency and the European Food Safety Authority, as follows:

David Bussard, United States Environmental Protection Agency, Washington, DC, USA  
(*Co-Chair*)

Weihsueh Chiu, United States Environmental Protection Agency, Washington, DC, USA  
(*lead author*)

Andy Hart, Risk and Numerical Sciences Team, The Food and Environment Research Agency, Sand Hutton, York, England, United Kingdom

Dale Hattis, Clark University, Marsh Institute, Worcester, MA, USA

Matthias Herzler, Federal Institute for Risk Assessment (BfR), Berlin, Germany (*lead author for Annex 5*)

Wout Slob, National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands (*lead author*)

Theo Vermeire, Expertise Centre for Substances, National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands (*Co-Chair*)

**Representatives**

Bernard Bottex, Unit of the Scientific Committee / SC Secretariat, European Food Safety Authority, Parma, Italy

George Fotakis, European Chemicals Agency, Helsinki, Finland

**Secretariat**

Kathy Hughes, Health Canada, Ottawa, Ontario, Canada

Carolyn Vickers, Chemical Safety, World Health Organization, Geneva, Switzerland

## LIST OF ABBREVIATIONS

ADI	acceptable daily intake
ADP	adenosine diphosphate
AF <sub>Inter</sub>	interspecies assessment factor
AF <sub>Inter-BS</sub>	interspecies assessment factor for body size differences
AF <sub>Inter-TK/TD</sub>	interspecies assessment factor for remaining toxicokinetic and toxicodynamic differences
AF <sub>Intra</sub>	intraspecies assessment factor
AF <sub>Intra-I</sub>	intraspecies variability assessment factor for target population incidence <i>I</i>
AF <sub>PoD-NOAEL</sub>	assessment factor for use of a no-observed-adverse-effect level as the point of departure
APROBA	approximate probabilistic analysis (tool)
AUC	area under the plasma or blood concentration versus time curve
BMD	benchmark dose
BMDL	lower confidence limit of the benchmark dose
BMDU	upper confidence limit of the benchmark dose
BMR	benchmark response
bw	body weight
C <sub>max</sub>	maximum concentration in blood or plasma achieved over time
CSAF	chemical-specific adjustment factor
CSAF <sub>Inter-TK</sub>	chemical-specific adjustment factor for interspecies toxicokinetics
DON	deoxynivalenol
EC <sub>50</sub>	median effective concentration
ECETOC	European Centre for Ecotoxicology and Toxicology of Chemicals
ED <sub>50</sub>	median effective dose
EEG	electroencephalographic
GLP	good laboratory practice
GM	geometric mean
GSD	geometric standard deviation
GSD <sub>H</sub>	geometric standard deviation for interindividual variability in the human equipotent dose distribution
GSD <sub>H-TD</sub>	geometric standard deviation for human interindividual toxicodynamic variability
GSD <sub>H-TK</sub>	geometric standard deviation for human interindividual toxicokinetic variability
GSD <sub>U</sub>	a measure of the uncertainty in the GSD <sub>H</sub>
HD	human dose
HD <sub>M</sub> <sup>I</sup>	target human dose; the human dose associated with a particular magnitude of effect <i>M</i> at a particular population incidence <i>I</i>
HED	human equivalent (oral) dose
<i>I</i>	population incidence
IC <sub>50</sub>	median inhibitory concentration
IPCS	International Programme on Chemical Safety

IPRA	integrated probabilistic risk assessment
$K_m$	Michaelis-Menten constant; the substrate concentration at which the initial reaction rate of an enzyme-catalysed reaction is half maximal
LC <sub>50</sub>	median lethal concentration
LCL	lower confidence limit
LD <sub>50</sub>	median lethal dose
LOAEL	lowest-observed-adverse-effect level
LVM	latent variable model
$M$	magnitude of effect
MOA	mode of action
NOAEL	no-observed-adverse-effect level
NTP	National Toxicology Program (USA)
OECD	Organisation for Economic Co-operation and Development
$P_x$	$x$ th percentile (e.g. P05, P50, P95)
PBTK	physiologically based toxicokinetics
PoD	point of departure
QSAR	quantitative structure–activity relationship
RBC	red blood cell
RDDR	regional deposited dose ratio
REACH	Registration, Evaluation, Authorisation and Restriction of Chemicals
RfD	reference dose
RGDR	regional gas dose ratio
RIVM	Rijkinstituut voor Volksgezondheid en Milieu (Dutch National Institute for Public Health and the Environment)
RtR	route-to-route
SD	standard deviation
SEM	standard error of the mean
TCA	trichloroacetic acid
TD	toxicodynamics
TDI	tolerable daily intake
TK	toxicokinetics
UCL	upper confidence limit
USA	United States of America
USEPA	United States Environmental Protection Agency
$V_{max}$	maximum initial rate of an enzyme-catalysed reaction
WHO	World Health Organization

## EXECUTIVE SUMMARY

### Purpose

This document provides guidance on quantitative approaches to evaluating and expressing uncertainty in hazard characterization for assessing the human health risk of chemicals, resulting in a harmonized approach to addressing the uncertainty in the final outcome of a hazard characterization. The guidance helps to make uncertainties associated with the outcome of any given hazard characterization more visible and allows for a more quantitative characterization of these uncertainties, so as to better inform risk management decisions. Specifically, it allows risk assessors to better communicate the range of possible risk implications of different risk management options to risk managers, while making the health protection goals associated with different options more explicit and transparent. The document also addresses uncertainties that are not easily quantified.

### Scope

The approach taken is an extension of existing approaches to hazard characterization. For instance, the identification of hazards, the approach to selecting critical studies and effects, and consideration of mode of action in determining human relevance are not affected by, and therefore are not discussed in, this guidance. The methodology provided focuses on addressing exposure to a single chemical for which dose–response information is available. Additional sources of uncertainty may arise from the use of alternative test methods (e.g. in vitro tests), the use of non-test methods (e.g. application of read-across) or other problem formulations (e.g. combined exposures to multiple chemicals). This document does not consider these situations, although the same general principles for evaluating and expressing uncertainty apply.

### Approach

Hazard characterization involves making inferences about the human population of interest for risk assessment (the “target population”) based on information (point of departure in the “study population”) on a specific end-point (“critical effect”) obtained from a scientific study. To evaluate and express uncertainty, a key conceptual distinction needs to be made between individual dose–response relationships, in which the magnitude of effect ( $M$ ) changes with dose, and population dose–response relationships, in which population incidence ( $I$ ) changes with dose. Therefore, the focus of the approach described in this guidance is to estimate the uncertainty in a “target human dose”, denoted  $HD_M^I$  and defined as the human dose at which a fraction  $I$  of the population experiences an effect of magnitude (or severity)  $M$  or greater (for the critical effect considered). A framework is therefore developed to evaluate and express uncertainty in the  $HD_M^I$ , making the choices for  $M$  and  $I$  explicit and transparent. Not only can the results of applying this framework be used to characterize the uncertainty associated with the output from more traditional approaches (e.g. dividing a point of departure by fixed factors), but, more importantly, the uncertainty range of the estimated  $HD_M^I$  itself can serve as the output of the hazard characterization, thereby providing better information to the risk manager compared with the single value provided in traditional approaches. Depending on the risk assessment needs as driven by the problem formulation, increasingly complex approaches may be employed.

## Framework

The framework outlined in this guidance includes the following basic steps:

1. Quantification of individual uncertainties (due to incomplete data or knowledge) in each aspect of the hazard characterization, including:
  - the point of departure;
  - any adjustments made due to characteristics of the study population or study design that differ from the target population or target conditions (e.g. interspecies differences, exposure duration); and
  - the amount of variability due to heterogeneity in the human population.
2. Combination of the uncertainties into the “overall” uncertainty of the final target human dose. Three increasingly complex approaches to combining uncertainties are presented in this document:
  - a non-probabilistic approach (multiplying upper and lower bounds);
  - an approximate probabilistic approach; and
  - a full probabilistic approach.

A simple, easy-to-use spreadsheet tool, “APROBA”, is provided for the execution of the approximate probabilistic approach.

3. Expression of the outcome in terms of ranges or probability distributions rather than as single (“conservative”) values.
4. If additional data or analysis may be needed to reduce uncertainties, estimates of the relative contributions from the various aspects to the overall uncertainties are used to identify the greatest sources of uncertainty. This information will show for which aspects additional information would be most effective in reducing the overall uncertainty.
5. Evaluation of uncertainties that are difficult to quantify, preferably by sensitivity analysis.

## Implementation and application

The guidance reviews available historical data to develop preliminary default uncertainty distributions, which can be used to implement this approach. The approach is illustrated by applying these distributions in a number of generic hazard characterization scenarios. Furthermore, the use of the results of the approach to answer some frequently encountered risk management questions is discussed. A detailed case-study applying the approach to a specific risk management scenario, using the APROBA spreadsheet tool, is provided.



## 1. INTRODUCTION

### 1.1 Purpose and scope

Hazard characterization, being part of risk assessment, aims to support decision-making. Depending on the particular decision, a hazard characterization can have multiple uses, such as to define a dose without appreciable health risk or to provide input for risk characterization of current exposures or of reduced exposures resulting from different actions. Ideally, the complexity of the hazard characterization will depend on the nature of the decision situation and the requirements for the precision of the outcome, but it will also depend on resources and the availability of data. In planning and scoping the risk assessment at the problem formulation stage, an idea about the decision options contemplated can guide the selection of the approach to the risk assessment.

The evaluation of uncertainties in risk assessment is a crucial issue, with direct consequences for risk management. The Codex Alimentarius Commission's working principles for risk analysis (Codex, 2013) state that:

uncertainties ... should be explicitly considered at each step in the risk assessment and documented in a transparent manner. Expression of uncertainty and variability in risk estimates may be qualitative or quantitative, but should be quantified to the extent that is scientifically achievable.

Risk managers need this information because, as also stated in Codex (2013), "the risk management options selected should reflect the degree of uncertainty and the characteristics of the hazard" and "the responsibility for resolving the impact of uncertainty on the risk management decision lies with the risk manager, not the risk assessors". Similarly, the United States Environmental Protection Agency's (USEPA) *Policy for risk characterization* (USEPA, 1995) states that:

Scientific uncertainty is a fact of life for the risk assessment process, and agency managers almost always must make decisions using assessments that are not as definitive in all important areas as would be desirable. They therefore need to understand the strengths and the limitations of each assessment and communicate this information to all participants and the public.

Consistent with this, USEPA criteria for transparency of risk characterization include "full disclosure of ... the major risk conclusions and the assessor's confidence and uncertainties in them" (USEPA, 2000). The USEPA (2000) guidance notes that "while it is generally preferred that quantitative uncertainty analyses are used in each risk characterization, there is no single recognized guidance that currently exists on how to conduct an uncertainty analysis". The present document contributes to filling that gap by providing guidance on quantitative approaches to evaluating and expressing uncertainty in hazard characterization.

Specifically, this document focuses on the question as to how uncertainties underlying a hazard characterization can be quantitatively evaluated and translated into an overall (again quantitative) statement on the uncertainty in the final outcome (e.g. a reference dose [RfD]). The approach discussed should be viewed as an extension of existing approaches of hazard characterization and should not be seen as interfering with existing approaches. For instance, the selection of the critical studies and effects, the principle of selecting the most sensitive end-points and the use of mode of action considerations in determining human relevance are not affected by, and therefore will not be discussed in, this guidance.

Whereas hazard identification addresses questions with “yes” or “no” answers, hazard characterization addresses questions that need a quantitative answer, such as a health-based guidance value. The evaluation of uncertainties in the first type of question has been recently addressed by others (Hart et al., 2010; EC, 2013; Edler et al., 2013). The evaluation of uncertainties in the second (quantitative) type of question is different, and this report is an attempt to provide a comprehensive discussion on that topic.

After the availability of relevant data is established, the specific problem formulation should translate into an overall approach of the hazard characterization to be done. At this stage, it should be clear which hazard characterization “aspects” are involved. An important aspect is dose–response assessment.<sup>7</sup> As dose–response information usually does not directly relate to the target population (defined in the problem formulation) and often does not directly relate to the target exposure situation, the outcomes from the dose–response assessment need to be extrapolated. Thus, the other aspects of hazard characterization relate to extrapolation of equipotent doses. Typical examples are adjusting the point of departure (PoD) for experimental animals into an equipotent human dose or adjusting the dose for typical humans into an equipotent dose for sensitive humans. By definition, extrapolation is associated with uncertainties, and a key issue in hazard characterization is how to deal with these uncertainties.

This document aims to provide guidance on how some important uncertainties in a given hazard characterization may be quantitatively evaluated, resulting in a harmonized way of expressing the uncertainty in the final outcome of the hazard characterization. It also provides some guidance on how to deal with uncertainties that are not easily quantified because of a lack of knowledge or data to inform the particular uncertainty. The methodology provided focuses on addressing exposure to a single chemical for which dose–response information is available. Additional sources of uncertainty may arise from the use of alternative test methods (e.g. in vitro tests), the use of non-test methods (e.g. application of read-across) or other problem formulations (e.g. combined exposures to multiple chemicals). This document does not consider these situations, but the general principles on quantifying uncertainty as discussed in this document apply in these cases too; the specification of the aspects involved would, however, need further work.

## 1.2 Use of quantitative evaluation of uncertainties in hazard characterization to achieve better informed and more transparent decisions

It has now been over half a century since the proposal of Lehman & Fitzhugh (1954) to represent uncertainties in projecting observed no-effect levels for toxic responses in small groups of animals to safe doses in the large population of humans by applying a combined factor of 100.

The flowering of biomedical sciences in the ensuing years has produced a wealth of information on interspecies differences in toxic responses, human diversity in sensitivity and several other uncertainties in utilizing toxicological data. Over the same period, our mathematical and computing technology has also taken giant steps towards increased

---

<sup>7</sup> This document focuses on situations for which experimental in vivo dose–response data are available for the chemical considered. Similar principles could be worked out for other situations, but these are not discussed here.

sophistication and capability. However, although the methodology has evolved to disaggregate the original 100-fold factor into individual subfactors representing specific aspects of hazard characterization, such as interspecies and intraspecies toxicokinetics (TK) and toxicodynamics (TD), these factors are still applied in a “deterministic” manner – that is, as point values that are combined by simple multiplication. This approach has some important drawbacks:

- The factors that are currently used are presumed to be “conservative”, in that they should “cover most cases”, which is defined only in a qualitative sense. Various terms have been used for these factors, such as safety factors (indicating that they are supposed to result in “safe” values), uncertainty factors (indicating that they are supposed to address the underlying uncertainties) and assessment factors (a more neutral term). Whatever terminology is used, there has been no specification of the degree of protection they are supposed to achieve, either individually or collectively, such as reduction of the incidence of adverse effects below X% in the population with Z% confidence. This implies that the outcome from a particular hazard characterization might be less or more conservative than would be considered desirable or necessary.
- Another drawback is that multiplying individual conservative values results in an overall factor that is even more conservative, the more so when the number of aspects in the specific hazard characterization is larger. This implies that the outcome from a particular hazard characterization might be more conservative than would be considered desirable or necessary.
- As a consequence of these issues, different hazard characterizations (related to different chemicals) may differ widely in the level of conservatism, depending on the specific values and the number of assessment factors that happened to be involved. This variation in the level of conservatism implies that the hazard assessments for different chemicals are not comparable, which complicates their use. For instance, it can result in risk management measures to focus on chemicals with the more conservative assessments, rather than those with the larger health risks.
- A further limitation of the current approach is that the outcome from the hazard characterization (e.g. an RfD) does not distinguish between the potency of the chemical and the uncertainties in the available information; both are merged into a single value.

These drawbacks can be addressed by characterizing the uncertainties in each aspect of the hazard characterization more formally and quantitatively, such as by statistical distributions, and by evaluating the overall uncertainty in the final outcome by probabilistic methods (Baird et al., 1996; Slob & Pieters, 1998; Vermeire et al., 1999; USEPA, 2001; Schneider et al., 2006; Van der Voet & Slob, 2007). In this way, both an upper bound and a lower bound for the hazard characterization outcome are obtained, making visible the overall uncertainty in that outcome. A further gain from multiplying distributions rather than single values is that the level of conservatism can be kept under control. Therefore, systematically applying probabilistic methods will allow assessors to develop assessments that are consistent and explicit in their level of conservatism, so that they are neither “too conservative” nor “not conservative enough” when compared with management-defined risk targets. Although Monte Carlo sampling is the most generally applicable approach for implementing probabilistic methods, a simple spreadsheet tool is being made available as an online companion to this guidance document that can implement an approximate probabilistic calculation, enabling the rapid application of probabilistic approaches.

Bringing this increased knowledge and analytical capability to bear on analyses of risks from exposure of humans to toxic chemicals has the promise to build an improved quantitative approach to analysis that will result in multiple benefits:

- More transparently representing in quantitative form the confidence we can have in toxicological risk projections and estimates of the relationship between dose and health effect, thereby facilitating choices of preventive measures and/or further information gathering by risk managers. For instance, health-based guidance values (e.g. the RfD) may be defined based on a pre-specified and harmonized level of conservatism, or estimated health risks for a given exposure situation can be expressed in terms of an uncertainty distribution or a confidence interval.
- Indicating which aspects contribute most to the overall uncertainty in the estimates will help direct limited risk assessment and risk management resources and, thus, guide the collection of better data and the creation of better analytical procedures to further improve risk analyses.

It cannot be known in advance how the improved quantitative information and analysis outlined in this document (and developed in future research) will affect decisions on measures to address toxic effects associated with environmental and occupational exposures to chemicals. That will be determined in the interactions among stakeholders that assign values to different policy propositions. However, it can be expected that better information will allow better informed and more transparent choices of how to respond to a wide variety of potential hazards.

### 1.3 Organization of this document

The remainder of this document is organized as follows:

- Section 2 gives an overview of this stepwise approach to evaluating uncertainty in hazard characterization and its relation to problem formulation.
- Section 3 discusses the basic concepts in detail.
- Section 4 shows how the basic inputs can be obtained for quantifying the uncertainty in the relevant aspects of a hazard characterization.
- Section 5 provides an illustration of the types of outputs that would result from probabilistic characterization of uncertainties.
- Section 6 discusses the interpretation and use of probabilistic outputs for some typical risk management questions.
- Section 7 proposes next steps with regard to research and implementation.

In addition, five annexes provide further details regarding some of the topics covered in this document:

- Annex 1 describes the three computational approaches to combining uncertainties introduced in sections 2 and 3.
- Annex 2 consists of a Microsoft Excel spreadsheet tool called “APROBA” (Approximate PROBabilistic Analysis) for making approximate probabilistic calculations of uncertainty, discussed in section 3, which is available online.<sup>8</sup>
- Annexes 3 and 4 provide details as to how historical data were used to develop the preliminary uncertainty distributions for common aspects of hazard characterization described in section 4. Annex 4 describes the data and analysis for intraspecies variability, whereas Annex 3 describes the data analysis for the other aspects.
- Annex 5 consists of a case-study illustrating the approach of probabilistic hazard characterization, using deoxynivalenol as the example compound.

---

<sup>8</sup> See the WHO/IPCS Harmonization Project publications webpage at: <http://www.who.int/ipcs/publications/methods/harmonization/en/>

## 2. CONTEXT AND OVERVIEW

### 2.1 Uncertainty in hazard characterization evaluated in the context of problem formulation for risk assessment

The importance of evaluating the uncertainty in the outcome from a hazard characterization can be recognized from the perspective of problem formulation for risk assessment. Problem formulation is a critical phase of the risk assessment process and provides the context for risk characterization.

As mentioned in other World Health Organization (WHO)/International Programme on Chemical Safety (IPCS) publications (e.g. Meek et al., 2013), problem formulation includes the risk management scope and goals in relation to relevant exposure scenarios, the level of uncertainty that is acceptable as well as the urgency of the assessment. The problem formulation can also include aspects such as characterization of experimental data sets and remaining uncertainty.

In analogy to the WHO/IPCS mode of action road map (Meek et al., 2013), uncertainty analysis, as shown in Fig. 2.1, is a process that is dependent on the problem formulation. Different options of uncertainty analysis can assist in the decision-making process for providing answers to the questions posed during problem formulation (e.g. establishment of health-based guidance values and protection goals).

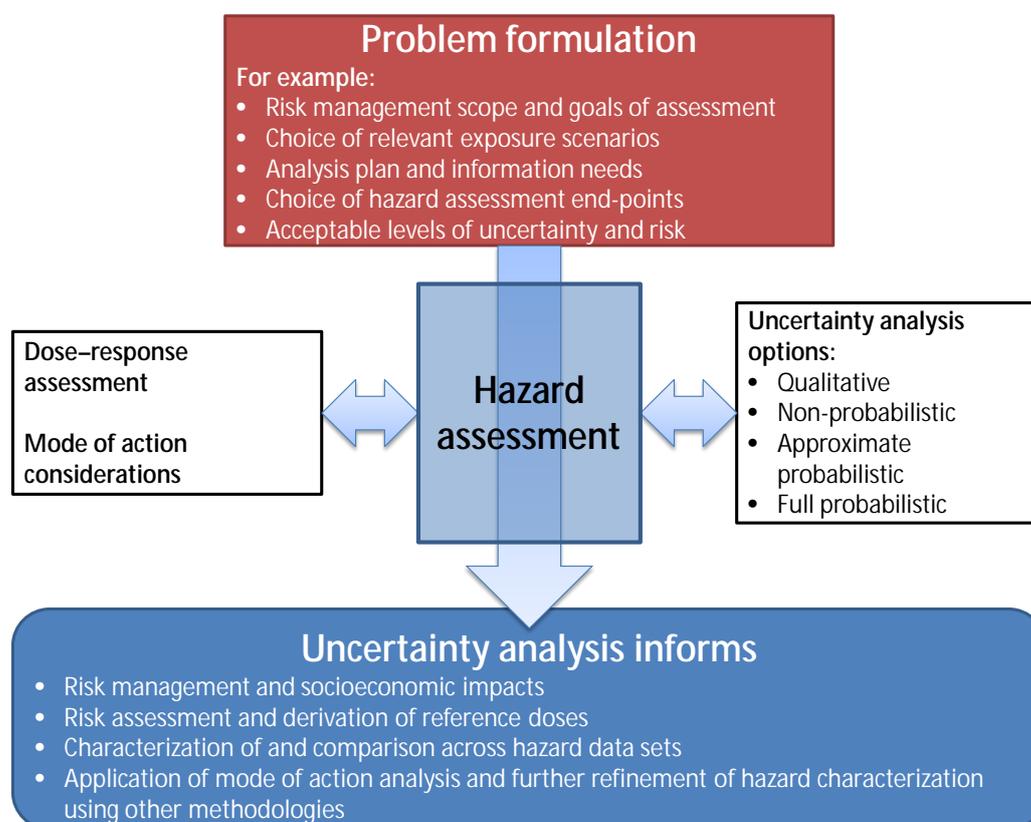


Fig. 2.1: Uncertainty analysis in the context of problem formulation

In addition, uncertainty analysis can apply to evaluating the added value of different levels of refinement of the assessment, as well as drive and inform the decision-making on what type of refinement may be preferable.

While it is logical that at least an initial formulation of the risk management problem should be the first step in the process of integrated risk assessment and risk management, this process in many cases is iterative. How much to invest in detailed exposure evaluation and analysis of risk management options will partially depend on initial assessments of the degree of current or forecasted exposure and risk. How much to invest in refined hazard characterization will in part be determined by consideration of the magnitude of current or expected exposure levels and how they compare with the results of the initial hazard characterization, as well as by the risk management framework being used.

## 2.2 Evaluation of uncertainty in hazard characterization in concert with other risk assessment methodologies

The process of uncertainty analysis based on problem formulation as sketched out in the previous section can inform risk assessors and risk managers in their analysis of experimental data sets and may also help improve application of the WHO/IPCS mode of action analysis framework for establishing human relevance (Meek et al., 2013) and other refinement methodologies, such as physiologically based toxicokinetic (PBTK) modelling, chemical-specific adjustment factors (CSAFs) and concordance analysis (IPCS, 2005, 2010; Meek et al., 2013), within risk assessment.

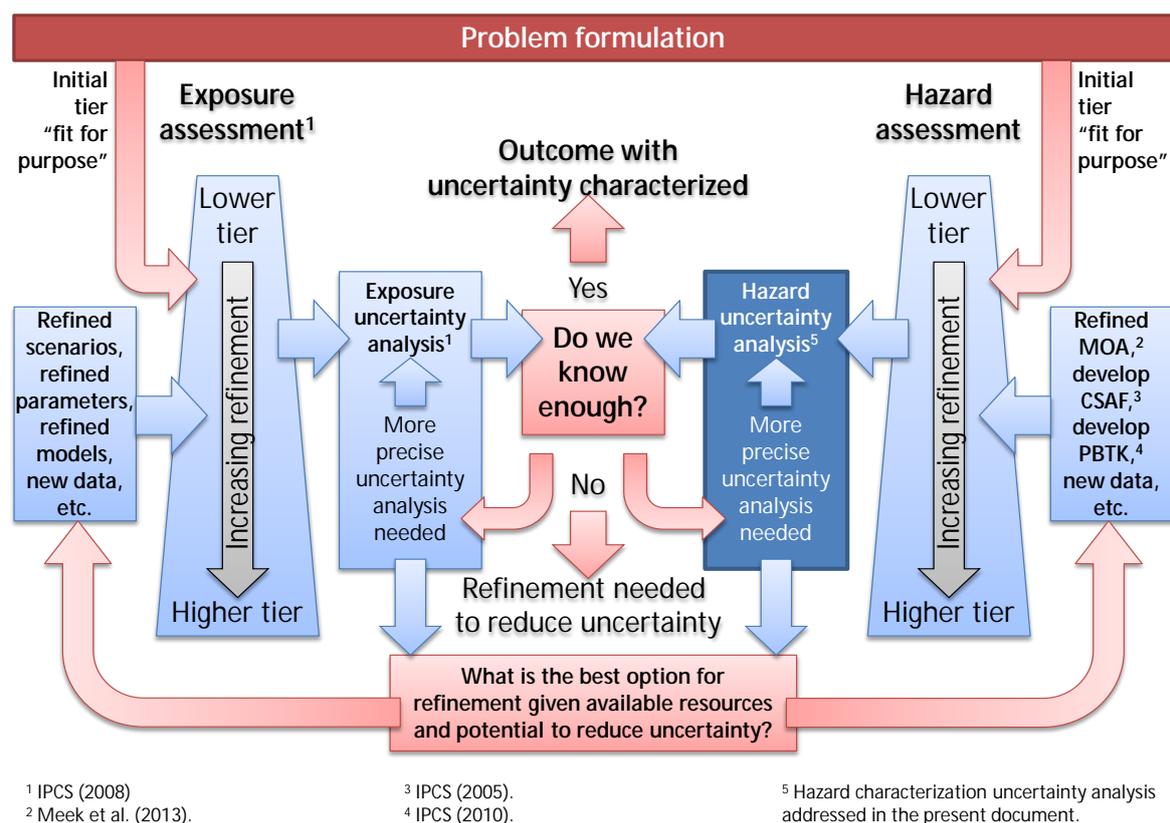
Depending on the risk assessment needs, a tiered approach can be applied to hazard assessment as well as to exposure assessment (see right- and left-hand sides of Fig. 2.2, respectively).

A tiering approach has also been described in the context of developing the WHO/IPCS combined exposures framework (Meek et al., 2011), with lower-level tiers dealing with default (conservative) assumptions for both exposure and hazard. Such tiering applies for either single-chemical or multiple-chemical risk assessments.

As shown in the right-hand side of Fig. 2.2 (the focus of this document), uncertainty analysis can be applied at any tier of hazard assessment. Higher uncertainty might be accepted for decisions and applications in relation to priority-setting exercises (lower tiers) compared with setting health-based guidance values (higher tiers).

Assessments made at lower tiers of hazard characterization can be refined by the incorporation of more data and more advanced models, including PBTK modelling and/or CSAFs. Higher tiers will result in better (more precise) estimates of health-based guidance values or of health risks at a given exposure level.

The basic idea behind a tiered approach is that at the end of a given tier in the assessment, the following question is raised: “Do we know enough?” (in the context of the problem formulation). If the answer is positive, the assessment can stop at that tier (and any remaining resources can be used for something else). If the answer is negative, a higher-tier assessment would be needed.



CSAF: chemical-specific adjustment factor; MOA: mode of action; PBTK: physiologically based toxicokinetic

**Fig. 2.2: A tiered approach in risk assessment including uncertainty analysis with reference to pertinent WHO/IPCS guidance**

Obviously, the question “Do we know enough?” is a question about uncertainties. More specifically, the question is: “Are the uncertainties acceptable, or are they not?” The answer is situation dependent and may, for example, require consideration of the exposure in addition to the hazard characterization outcome. For instance, the uncertainty in the outcome of a hazard characterization may be large, but if the lower confidence limit (LCL) of the outcome is far above the relevant exposure estimate, then the large uncertainty might be accepted without investing further resources, because one has confidence that current exposures do not pose a significant risk. If, however, the LCL of the outcome of the hazard characterization is lower than or close to the estimated human exposure and the uncertainty is considerable, there might be reason to go to a higher tier of hazard characterization to see if a refined analysis still supports a conclusion of potential risk.

A tiered approach to hazard characterization can lead to a more effective use of resources. However, a tiered approach will be more effective if the key question (“Do we know enough?”) is appropriately answered. This makes the evaluation of uncertainties a crucial part of hazard characterization; therefore, it is useful if uncertainties are quantified where possible. At the end of a hazard characterization, knowledge about the uncertainties in the final outcome can have a direct impact on the risk management decision to be made (see section 6).

This guidance does not intend to prescribe whether in a specific context it is appropriate to move to more refined exposure analysis, more refined hazard characterization or both. It does, however, show that a quantitative evaluation of the uncertainties can inform that decision.

## 2.3 Overview of the principles and framework for evaluating uncertainty in hazard characterization

In the process of developing a transparent approach for the quantitative evaluation of uncertainties in hazard characterization, a number of basic concepts arise that require further discussion. Most of these concepts are already in use, but need a more explicit definition in the form of a framework. In particular, implementing a quantitative uncertainty analysis forces a rethinking and explicit definition of the fundamental principles of the hazard characterization approach. The four fundamental principles underlying the framework for evaluating uncertainty in hazard characterization are summarized here and then further explained and discussed in section 3.

### **2.3.1 Principle 1: Individual-level effects (magnitude) and population-level effects (incidence) are conceptually distinct**

A conceptual distinction exists between individual dose–response relationships, in which the magnitude of effect  $M$  (often related to severity) changes with dose, and population dose–response relationships, in which the population incidence  $I$  of a particular magnitude of effect  $M$  changes with dose. Hazard characterization therefore focuses on the target human dose, or the human dose (HD) associated with a particular magnitude of effect  $M$  at a particular population incidence  $I$ , denoted as follows:

- ✚  $HD_M^I$ : the human dose at which a fraction  $I$  of the population shows an effect of magnitude (or severity)  $M$  or greater (for the critical effect considered).

### **2.3.2 Principle 2: For all types of end-points, the magnitude of effect $M$ can be regarded as changing gradually**

The magnitude of effect  $M$  is defined at the level of the individual. As discussed in more detail in section 3, all end-points can be represented in a form in which  $M$  gradually changes with dose.

Continuous dose–response data are explicitly measured as gradually changing (in each individual). For example, the measure of the per cent change in red blood cell counts is the continuous measure of magnitude in the following example of an  $HD_M^I$  for a continuous end-point:

- ✚  $HD_{05}^{01}$  (critical effect = red blood cell count): the human dose at which 1% of the population shows a decrease in red blood cell counts of 5% or greater.

Quantal data, where each individual is observed to respond or not for a specific effect, can also be conceptualized as resulting from an underlying continuous end-point that changes gradually with dose in an individual, as follows:

- In one case, the quantal data may be thought of as resulting from a process in which an underlying continuous end-point has a cut-point in how they are evaluated or reported (e.g. liver damage that can range in extent and other attributes may routinely be reported as to whether the liver lesion is of severity “mild” or greater). As dose increases, each individual experiences larger effects on a continuous scale, but the observation in the population is that more individuals have effects of at least a certain

size. Thus, an example for a deterministic quantal end-point would be that the “magnitude” of effect is “mild” in the  $HD_M^I$ :

✚  $HD_{mild}^{05}$  (critical effect = liver lesions): the human dose at which 5% of the population shows liver lesions of severity mild or greater.

- In another case, the quantal data may be thought of as resulting from a stochastic process in which each individual has a probability of a quantal effect, such as developing a tumour. The magnitude of the underlying probability of the quantal effect cannot be observed directly in an individual; as dose increases, however, each individual has a continuously increasing probability of experiencing the quantal effect, with the observation in the population that more individuals are observed to have the effect. Thus, in the following example of how  $HD_M^I$  would be interpreted for a stochastic quantal end-point, the “magnitude” of the effect would be a 5% extra risk of developing lung tumours:

✚  $HD_{05}^{01}$  (critical effect = extra risk of lung tumours): the human dose at which 1% of the population shows an individual extra risk of lung tumours of 5% or greater.

### **2.3.3 Principle 3: The concept of an “effect metric” for $M$ forms the basis of “equipotency” and differences in “sensitivity”**

The magnitude of effect  $M$  for which inferences are made is based on a selected “effect metric” defining “toxicological equivalence” or “equipotency”. This effect metric should measure the effect size in such a way that it applies across species (populations) as well as across individuals within a species (population). Changes of the same magnitude in this metric are considered to reflect equal toxicological changes (note that changes of equal size do not necessarily imply an equal level of adversity in different species or individuals; see discussion in section 3). Equipotent doses are therefore defined as doses that elicit the same magnitude  $M$  of the effect metric. Individuals with the same equipotent doses (at all effect sizes) are defined as equally sensitive to the chemical for the end-point. The key issue in defining effect metrics, discussed in section 3, is how to correct for differences in the background value of the particular parameter.

### **2.3.4 Principle 4: Making inferences from a point of departure involves making adjustments and accounting for variability and uncertainty**

Hazard characterization involves making inferences about the human population of interest for risk assessment (the “target population”) based on information obtained from a scientific study (the “study population”). In the usual deterministic approach, these inferences are accomplished using assessment factors to address (potential) differences due to differing species, human variability, suboptimal study conditions, etc. However, these factors are often mixtures of multiple elements that need to be clearly specified when evaluating uncertainties quantitatively. Specifically, making inferences between the “study” and “target” populations involves the following:

- making adjustments due to characteristics of the study population or study design that differ from the target population or target conditions;
- accounting for variability due to heterogeneity in the human population; and
- accounting for uncertainty associated with the two previous bullets due to incomplete data or knowledge.

For instance, there is some uncertainty as to exactly which factor to use to adjust for body size differences across species. Although there is considerable evidence to generally support scaling oral toxicity across species by an allometric power of body weight, there is some uncertainty as to the exact value of the allometric power. Similarly, variability within the human population needs to be taken into account in characterizing hazard, and the range of human responses is typically reflected with an estimate of intraspecies human variability. That estimate of human variability is again subject to uncertainty.

In some cases, chemical-specific information may be available that allows for more refined estimates of an assessment factor as applied to a chemical or class of chemicals. It is, however, important that such chemical-specific information (or information on a class of chemicals) not be used without full consideration of uncertainties in that information and in its implications for refining a more “generic” assessment factor.

### **2.3.5 Applying the framework**

The framework outlined by the above principles leads to the following basic approach to evaluating uncertainties:

1. Quantify individual uncertainties in each “aspect” of the hazard characterization – e.g. the uncertainty in the PoD – and in each adjustment or variability component that needs to be addressed (discussed in section 3.1).
2. Combine the uncertainties into the “overall” uncertainty of the final target human dose,  $HD_M^I$ . Three approaches to combining uncertainties are presented in this document, in the order of ascending precision in the estimates of overall uncertainty:
  - a. *Non-probabilistic approach*, where the individual lower and upper bounds for each hazard characterization aspect are combined by multiplication.
  - b. *Approximate probabilistic analysis*, where uncertainty distributions are combined probabilistically, assuming that all uncertainties can be described as independent lognormal probability distributions. The calculations may be performed without Monte Carlo simulations and are implemented in a spreadsheet tool, “APROBA”, which was developed with this document.
  - c. *Full probabilistic analysis*, where uncertainty distributions are combined probabilistically and are not restricted to independent lognormal probability distributions. Calculations are generally performed using Monte Carlo simulations.
3. Express the outcome in terms of ranges or probability distributions, rather than as single values.
4. If it is determined that additional data or analysis may be needed to reduce uncertainties, use the results of a probabilistic analysis to estimate the relative contributions from the various aspects to the overall uncertainty and thereby identify the greatest sources of uncertainty.
5. Evaluate remaining uncertainties that were considered difficult to quantify, preferably by sensitivity analysis.

An in-depth description of the framework has also been provided in the peer-reviewed literature (Chiu & Slob, 2015). Additional details regarding application of the framework are discussed in section 3.

## 2.4 Implications for problem formulation and interaction with risk management

As discussed above, the principles underlying the framework for evaluation of uncertainties in hazard characterization necessitate the introduction of the target human dose,  $HD_M^I$ , as the focus of analysis. The key implication of this framework for problem formulation and interactions with risk management is that it makes transparent that any dose estimated for hazard characterization reflects choices as to the magnitude of the effect  $M$  and the population incidence  $I$  of effects of that size or greater. Together with the quantified uncertainty (see section 2.4.2 below) in the estimated target human dose, this increases transparency by providing a better understanding of the outcome from the hazard characterization. It also allows risk assessors to better communicate the implications of value judgements in hazard characterization and the risk implications of different risk management options to risk managers. Conversely, it allows for problem formulation and the needs of the risk manager to be better communicated to the risk assessor. Some of these implications are explained in more detail below.

### **2.4.1 The target human dose $HD_M^I$ as the focus of hazard characterization when evaluating uncertainties**

#### *2.4.1.1 A conceptual transition to the $HD_M^I$*

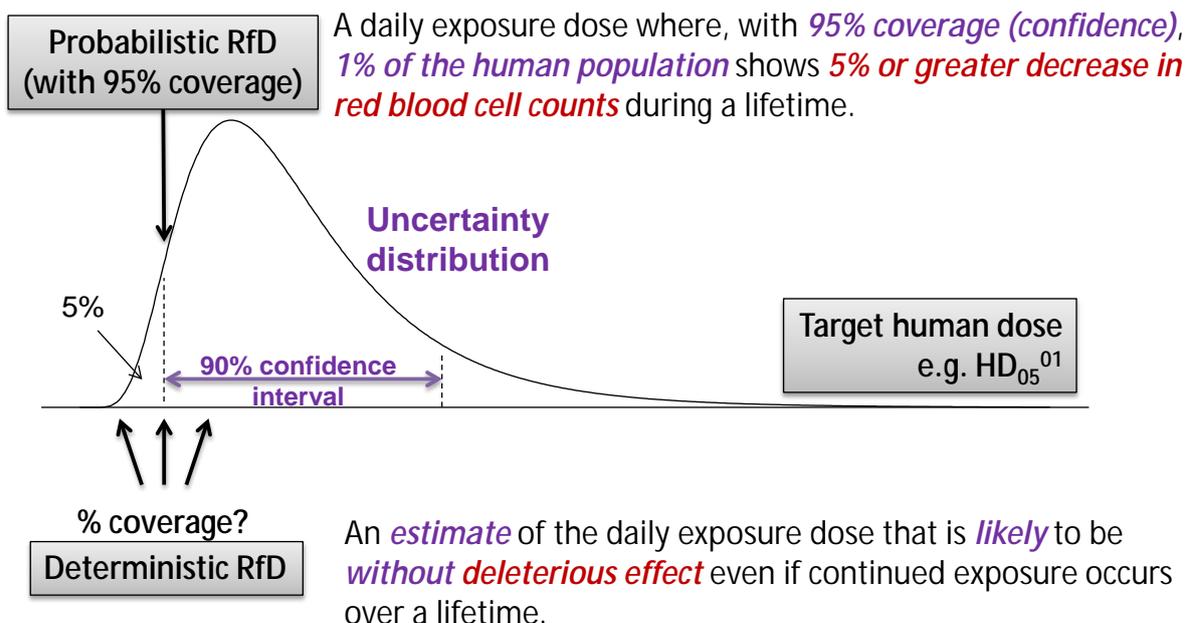
The fundamental principles described above for evaluating uncertainty in hazard characterization represent a shift from developing a health-based guidance value, such as a deterministic RfD, to estimating a target human dose, the  $HD_M^I$ . The deterministic RfD is calculated according to a given procedure, which is assumed to be conservative, but without saying how conservative in any individual case. The  $HD_M^I$  is defined as an unknown parameter that is estimated from the available information in the individual case. The precision in the estimated  $HD_M^I$  can be quantified in the form of an uncertainty distribution. This is illustrated in Fig. 2.3, contrasting the definitions of the deterministic RfD with those of a probabilistic RfD derived from the uncertainty distribution of the  $HD_M^I$ .

In setting a health-based guidance value such as an RfD, the  $HD_M^I$  distribution can be used in two ways, also illustrated in Fig. 2.3:

1. The  $HD_M^I$  uncertainty distribution can be used to estimate the “coverage” of a deterministic RfD – that is, the per cent confidence that this RfD protects the population against a specified magnitude and incidence of effect.
2. Additionally, the LCL of the  $HD_M^I$  can be used as a probabilistic RfD to replace the deterministic RfD. In this case, the probabilistic RfD is the dose that protects the population from a specified magnitude and incidence of effect with a pre-specified per cent coverage (confidence).

#### *2.4.1.2 Transition from the deterministic RfD to the $HD_M^I$ as a natural extension of the transition from the NOAEL to the BMD*

In understanding how the RfD and  $HD_M^I$  are related to each other, it may be helpful to review the relationship between the no-observed-adverse-effect level (NOAEL) and the benchmark dose (BMD), as shown in panel A of Fig. 2.4. In particular, the description of a NOAEL does not precisely specify the level of response, other than that it is below the response level that



Target human dose	Probabilistic RfD	Deterministic RfD
HD	"daily exposure dose"	"daily exposure dose"
$M$ (magnitude)=5%	"5% or greater decrease in red blood cell counts"	"deleterious effect"
$I$ (incidence)=1%	"1% of the human population"	"likely...without" (in the sense of variability)
Uncertainty	"95% coverage (confidence)"	"estimate...likely" (in the sense of uncertainty)

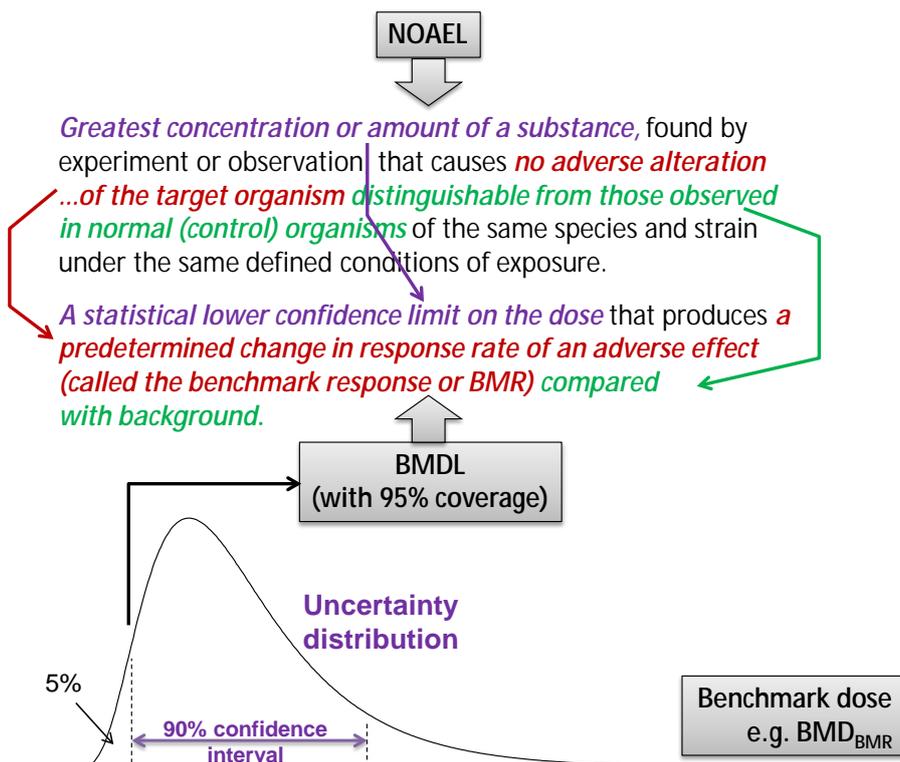
**Fig. 2.3: Illustration of the contrast between the definitions related to the deterministic RfD and those related to a probabilistic RfD derived from the uncertainty distribution of the  $HD_M^I$ .** Note that in the deterministic RfD, the term "likely" can be interpreted as relating to both variability (i.e. "most people" are "without deleterious effect") and uncertainty (i.e. is the statement actually true?).

can be observed in the utilized study. It is sometimes assumed to be a level associated with "no adverse effect". In fact, the NOAEL is often identified based on lack of statistical significance, which depends strongly on the study design (number and spacing of dose levels and number of animals per dose). Additionally, because it has to be one of the dose levels reported, the uncertainty in the NOAEL is considerable, but cannot be quantified within this approach. In contrast, the BMD requires precise specification of the level of response – the benchmark response, or BMR. The BMD is identified by interpolation from the dose–response relationship, and aspects of study design, such as applied doses and number of animals per dose group, are used to quantify the uncertainty in the BMD.

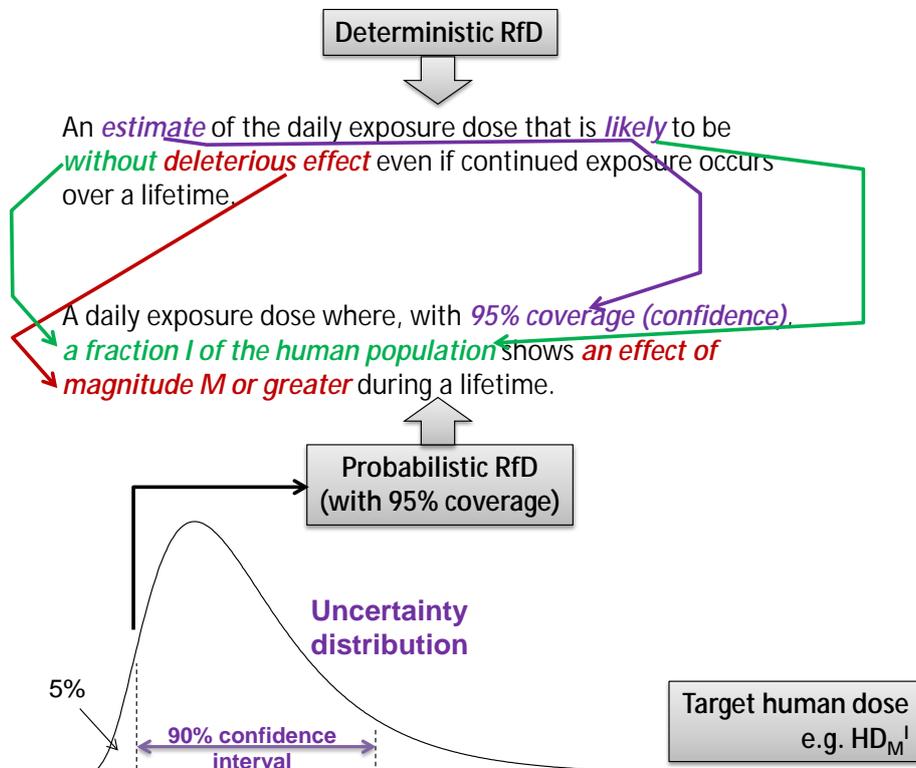
Similarly, as shown in panel B of Fig. 2.4, the deterministic RfD does not precisely specify:

- the individual level of response, other than that it is sometimes assumed to be "without deleterious effect";
- the incidence of response at the population level, other than that it be "likely" in the context of variability (i.e. "most people" are "without deleterious effect"); or

A



B



BMD: benchmark dose; BMDL: lower confidence limit of the benchmark dose; BMR: benchmark response;  $HD_M^I$ : human dose associated with a particular magnitude of effect  $M$  at a particular population incidence  $I$ ; NOAEL: no-observed-adverse-effect level; RfD: reference dose

**Fig. 2.4: A. Comparison of the NOAEL with the BMDL; B. Comparison of the deterministic RfD with the probabilistic RfD based on the  $HD_M^I$**

- the uncertainty, other than that it be an “estimate” that is “likely” to be correct in the context of uncertainty (i.e. uncertainties in the assessment have been adequately covered).

In contrast, the  $HD_M^I$  requires quantitative specification or estimation of all three of the following:

1. the magnitude of effect  $M$  at the individual level;
2. the population incidence  $I$ ; and
3. the uncertainty in the form of a probability distribution.

Just as it is the BMDL – the LCL of the BMD – that serves as the replacement for the NOAEL, it would be the LCL of the  $HD_M^I$  (the probabilistic RfD) that would serve as a replacement for the deterministic RfD.

In summary, just as the BMDL represents a PoD that is more precisely defined quantitatively than the NOAEL, a probabilistic RfD derived from the  $HD_M^I$  represents a health-based guidance value that is more precisely defined quantitatively than a deterministic RfD.

#### ***2.4.2 Being “conservative” and its relationship to evaluating uncertainty***

Roughly speaking, the term “conservative” is used in the sense of aiming to be “on the safe side”. It plays a central role in risk management and hazard characterization, but a closer look reveals that the term is used with different meanings in different contexts. The framework described in this document, and in particular the introduction of the target human dose,  $HD_M^I$ , allows for more precision to be used in its definition.

##### *2.4.2.1 “Conservative” in relation to protection goals*

One use of the term conservative could be in relation to having strict protection goals, such as a low value for the acceptable magnitude of effect  $M$  or a low value for an acceptable incidence  $I$  given the seriousness of the effect. Similarly, in exposure assessment, one may focus on a highly exposed individual in the population and thus be conservative with respect to characterizing the exposure of the typical individual to the chemical. The term conservative may also be used in terms of specifying a sensitive “target population”, such as individuals who are vulnerable due to a poor health status (e.g. asthma patients in a context of inhalation exposure; the concept of “vulnerability” is further discussed in section 3.2.3 below).

##### *2.4.2.2 “Conservative” in relation to uncertainty: “coverage” and “degree of uncertainty”*

Being conservative can also take place in the context of uncertainty. For this situation, there are two possible interpretations, which may be denoted by the terms “coverage” and “degree of uncertainty” and are best explained using examples.

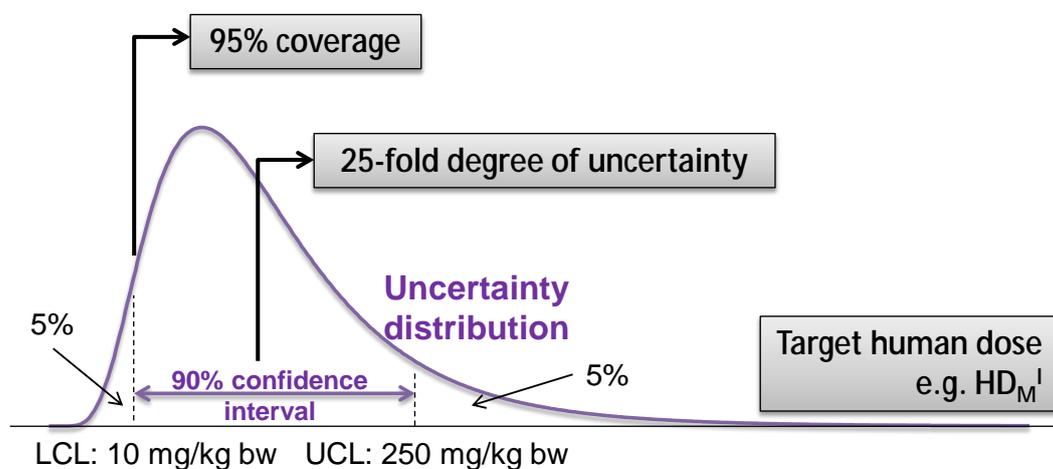
For instance, based on historical data, there is a distribution of ratios of the subchronic PoD (i.e. the starting point for hazard characterization, such as a NOAEL or BMDL) to the chronic PoD for a large number of chemicals. While this distribution reflects variability among chemicals, it can be interpreted as an uncertainty distribution for an individual chemical in hand for which no chronic study is available. In that case, a conservative value for the subchronic assessment factor may be obtained by choosing a high percentile of the

distribution of these ratios of the subchronic to chronic PoDs. Suppose a factor of 10 is chosen, and 99% of the chemicals have a ratio lower than that. Then one could state that the factor of 10 “covers” 99% of the chemicals (i.e. its coverage is 99%).

For the same distribution of ratios of subchronic to chronic PoDs, we may consider the distance (ratio) between a higher and a lower percentile (e.g. 95th percentile [P95] to 5th percentile [P05]). When this distance is large, there is more uncertainty in extrapolating a subchronic PoD to a chronic PoD than when this distance is small. Therefore, this distance measures the degree of uncertainty in extrapolating a subchronic to a chronic PoD.

Similarly, given a probabilistic hazard characterization that results in an estimate of the target human dose in terms of the confidence interval – (10, 250) mg/kg body weight (bw) – with 90% confidence (two-sided), as shown in Fig. 2.5, the “coverage” and “degree of uncertainty” can be illustrated as follows:

- When the LCL of 10 mg/kg bw is selected as a potential health-based guidance value, then its coverage is 95%, which may be interpreted as follows: there is a 95% probability that the (true) target human dose is higher than 10 mg/kg bw.
- The degree of uncertainty in this example is a factor of 25, meaning that the target human dose might be up to a factor of 25 higher (with a 5% probability that it is more than a factor of 25) than the lower end of the confidence interval. For a given level of “conservatism”, the size of this ratio from the upper to the lower end of the confidence interval is a measure of how much uncertainty there is in the estimate.



bw: body weight;  $HD_M^I$ : human dose associated with a particular magnitude of effect  $M$  at a particular population incidence  $I$ ; LCL: lower confidence limit; UCL: upper confidence limit

**Fig. 2.5: Illustration of the concepts of “coverage” and “degree of uncertainty”, given an uncertainty distribution for the  $HD_M^I$**

Both measures of conservatism are relevant in the context of risk management. When the LCL of the target human dose distribution is used as an RfD, then its associated degree of conservatism can be expressed either in terms of the associated coverage or in terms of the degree of uncertainty – that is, the phrase “more conservative” can mean either a higher coverage or a larger degree of uncertainty (or vice versa). Note, however, that the coverage can be chosen at will, whereas the degree of uncertainty is an outcome of the assessment related to the information available in that specific case.

One of the main messages of this document is that risk assessors should inform risk managers by providing not only the conservative risk estimate (upper bound for risk or lower bound for

target human dose), but preferably also the non-conservative estimate (lower bound for risk or upper bound for target human dose). Even in a context where the non-conservative value will not by itself be used to set a regulatory target, it helps communicate the range of doses estimated to result in the target incidence and magnitude of effect and the degree of uncertainty associated with the conservative value. As an illustration, imagine two risk assessments (for two chemicals) that result in the same upper bound for the predicted risk. For a risk manager, if human exposures to these chemicals were similar and were above that value, this could imply that both chemicals would need equal attention. However, if the lower-bound risk estimate is much higher for the first chemical than for the second chemical, then this additional information might be a reason for the risk manager to prioritize the first chemical over the second based on the available information or to prioritize the second with respect to seeking additional data. The relationship between both aspects of uncertainty (conservatism) and various risk management issues will be further discussed in section 6.

### **2.4.3 Using conclusions from analysis of uncertainties to inform the question, “Do we know enough?”**

As discussed in section 2.2, and as shown in Fig. 2.2, uncertainty analysis feeds directly into the key question, “Do we know enough?”, which is a central decision point in the tiered approach to risk assessment. More specifically, this question is answered by making a judgement as to whether the uncertainty, in terms of both “coverage” and “degree of uncertainty,” is acceptable or not – a judgement that will depend on the specific case, including the type of effect, the specific problem formulation and the consequences of potential risk management decisions or measures. Consider, as an example, a health-based guidance value (e.g. RfD) derived in the traditional way. From a human health perspective, a high per cent coverage would be desirable. However, if the estimated or expected human exposure were substantially lower than this RfD, a high coverage may be less critical. As another example, if the coverage is sufficient but the degree of uncertainty is very large, the RfD might be much smaller than would prove to be needed if the uncertainty could be reduced. In situations where there would be substantial socioeconomic consequences from accepting this RfD, it may not be desirable to accept such large uncertainty, and therefore it may be worthwhile to spend further resources to try to reduce it.

The results of an uncertainty characterization (at any tier of the hazard characterization) may lead to one of the following conclusions:

1. The uncertainty is considered acceptable in terms of both coverage and degree of uncertainty in the context of its intended use for hazard characterization. Therefore, the assessment may stop here.
2. The uncertainty is not considered acceptable in terms of either coverage or degree of uncertainty in the context of its intended use for hazard characterization.
  - a) If coverage is not acceptable, an alternative value can be used that has the desired coverage; this is easily calculated using the APROBA tool.
  - b) If the degree of uncertainty is unacceptably large, then a higher-tier hazard characterization may be able to reduce the degree of uncertainty – for example, by generating new data.
  - c) A sensitivity analysis may be useful before deciding how to spend resources for additional information. In particular, a probabilistic analysis (if not already done) can estimate the relative contributions from the various aspects to the overall uncertainty. The information from this sort of sensitivity analysis may assist in

targeting further analysis or data generation in line with the largest sources of uncertainty.

3. The uncertainty is felt to be inadequately characterized in the context of its intended use for hazard characterization. In that case, a more advanced probabilistic analysis is called for (i.e. an approximate probabilistic analysis to replace a non-probabilistic analysis, or a full probabilistic analysis to replace an approximate probabilistic analysis).
4. The uncertainty is either unacceptable or inadequately characterized, but there are inadequate resources or time to conduct a more refined assessment. In those cases, the risk management decision needs to be taken in the face of uncertainty. This uncertainty has now been made visible in a quantitative way, however, which makes the risk management decision better founded and more transparent than a decision based on results for which the uncertainty is unknown. The uncertainty analysis could at the same time inform plans to obtain further data or do additional analysis for a future revision of a decision that had to be made based on the currently available data and analysis.

In the cases where a higher-tier hazard characterization is called for, the current uncertainty could be reduced in various ways. For instance, when the uncertainty is judged unacceptable in an assessment that was based on a NOAEL, the first option to consider is applying the BMD approach and estimating the BMD uncertainty distribution to replace the NOAEL, as that may reduce the overall uncertainty, as will be illustrated in section 6 and the case-study in Annex 5. Additional options include generating additional dose–response data, developing a CSAF (with an associated uncertainty distribution) or using PBTK modelling (in which uncertainties are taken into account). When this higher-tier assessment is completed, the question of “Do we know enough?” is reconsidered, so that the process iterates until the conclusion “Stop here” is reached. After a higher-tier assessment has been performed, updating the probabilistic analysis also makes visible to what extent the additional data or models were effective in the sense of reducing the overall uncertainty in the final outcome of the hazard characterization.

It is reiterated that the answer to the question “Do we know enough?” strongly depends on the specific case, in a way that cannot easily be captured in generic terms. Section 6 provides a number of example cases, which illustrate how the decision to require a higher-tier assessment or not may be made.

#### **2.4.4 How to adapt the approach for different regulatory contexts**

Many of the examples in this document illustrate how the estimation of the  $HD_M^1$  might be used to establish a health-based guidance value or to determine if any risk management action is needed (by examining whether current exposures exceed the uncertainty range of the target human dose,  $HD_M^1$ ). The prior discussion also illustrates how the explicit information about uncertainty can be used to inform decisions, not just about tiering of the analysis and for investments in research, but also for setting priorities for action.

The approach presented in this document can also be adapted for use in other regulatory contexts. For example, the approach can be utilized to illustrate for a risk manager how different target exposures would relate to different values of the target incidence for a given effect or different magnitudes of effect. This could help risk managers balance the options for reduction in human health risk against other social or economic impacts of risk management actions. Examples of how to illustrate the relationship between different incidence levels are provided later in this document.

### 3. DETAILED DESCRIPTION OF THE APPROACH TO EVALUATING UNCERTAINTY IN HAZARD CHARACTERIZATION

#### 3.1 Aspects of hazard characterization and associated uncertainties

Hazard characterization is a process that uses a PoD established in a study population (usually experimental animals, sometimes humans) with the purpose of estimating an equipotent dose in the target population (usually sensitive humans). In this process, a number of discrete quantitative aspects can be distinguished, beginning with analysis of the dose–response data and then making a number of adjustments to arrive at the final quantitative output. Table 3.1 provides a summary of possible aspects, all of which are subject to uncertainties. In principle, the uncertainties for each of these aspects can be quantified based on historical data. Section 4 illustrates this for a number of aspects (those labelled by the superscript “a” in Table 3.1).

**Table 3.1: Common aspects of hazard characterization**

<i>Aspect</i>	<i>Description</i>
Determination of a PoD <sup>a</sup>	There are two (statistical) approaches to quantify the PoD: the NOAEL approach and the BMD approach. Although both approaches differ in many ways, they basically have the same purpose: to estimate a dose at which the effect is small. Thus, at the NOAEL, the effect is assumed to be of a small (given the group sizes used and the supposedly sufficient power of the statistical test), but unspecified, magnitude. The uncertainty in the NOAEL, although not usually considered in current practice, is not negligible and can be taken into account (see section 4). In the BMD approach, the BMDL is a dose at which the effect is likely (in a statistical sense) to be smaller than an explicitly specified effect size, so uncertainty is explicitly taken into account.
Interspecies extrapolation <sup>a</sup>	The purpose of this aspect is to make inferences of toxicity in humans based on a toxicity study in a different species. Specifically, this aspect estimates the equipotent dose between the typical human being and the test animal. It has two components: adjusting the dose for generic physiological differences (e.g. divide by body weight or use allometric scaling for oral doses; respiratory tract differences for inhalation concentrations) and accounting for potential differences in sensitivity between test species and humans for the specific chemical considered. The latter may be due to chemical-specific TK/TD differences between species. When the appropriate data are available, a CSAF may be derived, possibly assisted by mathematical models (IPCS, 2005).
Estimating intraspecies variability <sup>a</sup>	The purpose of this aspect is to account for variations in sensitivity within the human population, in terms of the ratio of equipotent doses when comparing typical with “sensitive” human beings. Traditionally, the sensitive human being is not defined in a quantitative sense. In some cases, a chemical-specific intraspecies factor may be derived, possibly assisted by models (e.g. population PBTK models).
Extrapolating across dosing duration <sup>a</sup>	The purpose of this aspect is to make inferences on effects in a population exposed for a different (usually longer) duration than in the toxicity study. Specifically, this aspect estimates the ratio of equipotent doses for different durations of exposure (e.g. the chronic dose eliciting the same magnitude of effect as the equipotent subchronic dose).

**Table 3.1 (continued)**

<b>Aspect</b>	<b>Description</b>
Extrapolating across dosing patterns	The purpose of this aspect is to make inferences as to toxicity from the exposure pattern in the population of interest based on a different exposure pattern in an available toxicity study. For example, exposure in the target population might be continuous, whereas the test animals were exposed for 8 hours per day or for 8 hours per week. This case is usually covered by a proportional correction factor. As another example, humans may be exposed by (irregular) peaks, whereas the animals received constant doses over time. Such situations are difficult to handle, and no generally accepted approach exists.
Extrapolating to low-effect levels <sup>b</sup>	The purpose of this aspect is to make inferences about doses associated with lower levels of effect than observable (or observed) in the toxicity study. Specifically, this aspect estimates the dose eliciting a particular magnitude of effect given a dose associated with a higher magnitude of effect. One example is linear extrapolation, in which it is assumed that risk decreases proportionally with dose. Another example is LOAEL to NOAEL extrapolation, which is often done by applying an arbitrary assessment factor, usually 10 or 3.
Estimating the impact of missing studies (key end-points, dose levels)	The purpose of this aspect is to make inferences as to the potential impact of missing toxicity studies. Specifically, given the results based on a particular study and end-point, this aspect estimates how much lower the resulting dose would be if based on the most sensitive end-point from a database including additional, currently missing, toxicity studies (e.g. missing a developmental study).
Extrapolating across agents	The purpose of this aspect is to make inferences when there are only data on similar or related chemicals or chemicals thought to act by similar ways, rather than the chemical of interest. Specifically, given the results based on similar or related chemicals, this aspect estimates what the dose associated with a similar end-point and magnitude of effect would be for the chemical of interest. Read-across, in which end-point information on one chemical is used to predict the dose–response relationship for the same end-point in another (comparable) chemical, is an example. Chemical categories based on a group of chemically similar substances may allow for an analysis of trends in certain toxic properties, but this interpolation may also be associated with substantial uncertainty. QSARs are based on a large number of chemicals, relating equipotent doses to a quantitative characteristic of the chemicals in the test set, such as binding affinity for a particular receptor. Uncertainty analysis here has to address questions such as the suitability of the model to successfully predict the target end-point or the accuracy with which its domain of applicability can be characterized, but also substance-specific aspects, such as to what extent the target chemical falls into that applicability domain.
Extrapolating across exposure metrics	The purpose of this aspect is to address the situation in which the measure of exposure or dose differs between the population of interest and the population providing the dose–response information. This includes the relationship between external exposure and internal dose metrics (e.g. lead in drinking-water to lead in blood), as well as the relationship among different internal dose metrics (e.g. lead in bone to lead in blood). This may utilize empirical data on the relationship between metrics or a TK (e.g. PBTK) model.
Extrapolating from early to late effect	The purpose of this aspect is to address the situation in which there are data on early (“upstream”) biological effects and one wants to predict later (“downstream” or “apical”) effects, such as disease end-points. Examples might be extrapolating from blood pressure to stroke or from reduced birth weight to neonatal mortality.
Extrapolating from in vitro or in chemico to in vivo data	The purpose of this aspect is to make inferences on in vivo effects from in vitro data. Examples are the use of in vitro tests for testing the irritating or corrosive potential of chemicals or corrosivity or the determination of sensitizing potential based on chemical reactions with certain protein functional groups in chemico.

Table 3.1 (continued)

BMD: benchmark dose; BMDL: lower confidence limit of the benchmark dose; CSAF: chemical-specific adjustment factor; LOAEL: lowest-observed-adverse-effect level; NOAEL: no-observed-adverse-effect level; PBTK: physiologically based toxicokinetic; PoD: point of departure; QSAR: quantitative structure–activity relationship; TK/TD: toxicokinetic/toxicodynamic

<sup>a</sup> Uncertainties in these aspects are discussed and quantified on a preliminary basis in section 4.

<sup>b</sup> Uncertainties in these aspects are discussed, but not quantified, in section 4.

### 3.2 A framework for evaluating uncertainty in hazard characterization based on four fundamental principles

As summarized in section 2, the framework for evaluating uncertainty in hazard characterization developed in this document rests on four fundamental principles. These principles are discussed here in more detail.

#### **3.2.1 Principle 1: Individual-level effects (magnitude) and population-level effects (incidence) are conceptually distinct**

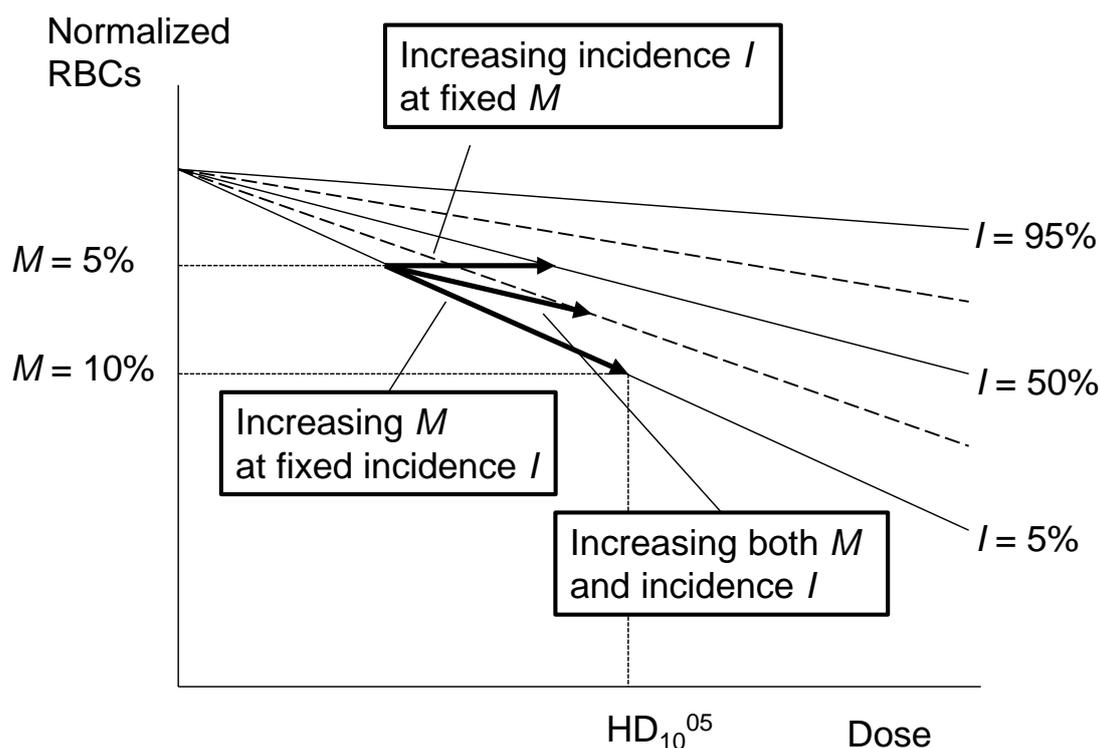
The starting point of this framework is that a conceptual distinction exists between effects on individuals and effects on the population. In particular, the effect of exposure at the level of the individual is a change in the magnitude of a measure of toxicological effect. The effect of (a fixed) exposure in a population is a change in the magnitude of effects in each individual, resulting in a change in the incidence of effects of any particular magnitude. In the present approach, the magnitude of change needs to be ordinality related to severity – so a greater magnitude constitutes a more severe effect. For instance, a body weight decrease of 20% is greater in magnitude (and is more severe) than a body weight decrease of 10%, and a “moderate” liver lesion is greater in magnitude (and is more severe) than a “mild” liver lesion. Thus, for a monotonic dose–response relationship in an individual, it may be imagined that a higher exposure will, for any given end-point, lead to effects that are larger in magnitude (and hence greater in severity). In a human population, increasing exposure levels will result in a higher incidence of individuals at or above a given magnitude of effect for the end-point considered. It will also result in increasing magnitudes of the effect for a fixed percentile of the population. Thus, as magnitude of effect and incidence related to a given end-point increase at the same time, more and more subjects will suffer from more and more severe effects (i.e. of larger magnitude) as exposure increases (see Fig. 3.1 for an illustration).

In order to explicitly and quantitatively evaluate uncertainties, the distinction between magnitude (or severity) and incidence needs to be explicitly maintained in a hazard (or risk) characterization. For example, when the aim is to derive a human limit value, then the associated<sup>9</sup> target human dose is defined as a function of both the magnitude of the effect and the fraction of the population with that effect. For convenience, we establish the notation whereby human dose or exposure is denoted HD, the magnitude of effect is denoted by  $M$ , incidence is denoted  $I$  and their combined relationship is denoted as follows:

  $HD_M^I$ : the human dose at which a fraction  $I$  of the population shows an effect of magnitude (or severity)  $M$  or greater (for the critical end-point considered).

---

<sup>9</sup> Note that a health-based guidance value derived in a hazard characterization would not be the same as the target human dose, but rather would be a conservative estimate of it.



**Fig. 3.1: Increasing magnitude of the effect ( $M$ ) and incidence ( $I$ ) with dose.** The middle line reflects the hypothetical dose–response relationship for a decrease in red blood cells (RBCs) in the median individual (hence,  $I = 50\%$ ), the lowest line that of a more sensitive individual (at the 5th percentile [P05] of the population), and the topmost line that of a less sensitive individual (95th percentile [P95]). The dose–response relationships are normalized to each individual's own background value on the y-axis. For a given effect size (e.g.  $M = 5\%$  decrease in RBCs), a higher dose will result in a higher incidence (see shortest arrow). For a given percentile of the population (e.g.  $I = 5\%$ ), a higher dose will be associated with a larger effect size  $M$  (see longest arrow). Similarly, a higher dose can also be associated with a simultaneous increase in  $I$  and  $M$  (see middle arrow).  $HD_{10}^{05}$  represents the human dose (HD) at which a 10% (or greater) magnitude of effect ( $M$ ) is experienced at a population incidence ( $I$ ) of 5%, a notation that is explained in the text.

This notation indicates the (estimated) human dose with the specified magnitude of effect  $M$  and incidence  $I$  given that magnitude of effect. A major advance of this framework is the specification of  $HD_M^I$  as the final goal of hazard characterization; in the past, the distinction between severity and incidence has usually not been made explicit. Specification of the value of  $M$  for different types of end-points is discussed in the next two sections.

### 3.2.2 Principle 2: For all types of end-points, the magnitude of effect $M$ can be regarded as changing gradually

Toxicological dose–response data may relate to continuous end-points or to quantal (categorical) end-points. Specifying the magnitude of effect  $M$  for these different types of end-points is discussed below.

#### 3.2.2.1 Continuous end-points

Changes in continuous end-points reflect changes in severity (in the sense of magnitude of the effect), which can be imagined to increase with dose in each individual. In most cases, the changes in severity with dose cannot be directly measured in each individual, as individuals receive only one dose in most studies. However, they can be inferred from changes in groups exposed to different dose levels. Therefore, for continuous end-points, the magnitude of

effect takes the form of a quantified level (e.g. a per cent change in the biological parameter). An example of this in the context of specifying  $HD_M^I$  is:

- ✚  $HD_{05}^{03}$  (critical effect = red blood cell count): the human dose at which the decrease in red blood cell count is 5% (or greater) in 3% of the population.

Thus, 97% of the population would be protected from a 5% (or higher) decrease in red blood cell count. Holding severity  $M$  constant, lower doses (lower values of HD) will result in a lower fraction  $I$  of the population experiencing an effect of severity  $M$  or greater. Holding  $I$  constant, lower doses will result in a lower severity of effect, such as a 2% rather than a 5% decrease in red blood cell count, in this fraction of the population.

### 3.2.2.2 Quantal end-points

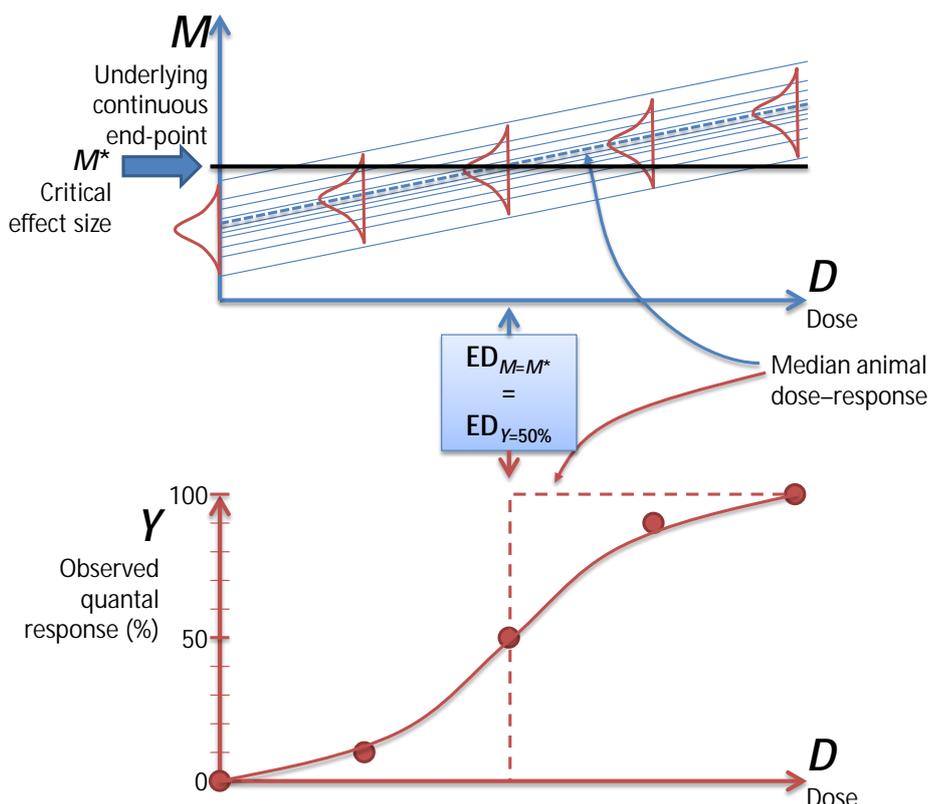
For quantal end-points, such as histopathological changes, the change in magnitude of effect is often measured in terms of severity categories, such as minimal, mild, moderate and severe. These are reflections of an underlying gradually changing severity, which is often difficult to capture in a quantitative measure. Thus, dose–response data on effects such as histopathological changes may be reported as incidences for each severity category (at each dose). However, they are often reported (or analysed) as quantal dose–response data, with the incidences relating to a single severity category – for example, the number of animals with at least mild (or other severity category) liver lesions.

Therefore, for quantal data on, for example, histopathological end-points, the dose–response relationship reflects the increase in the observed fraction of affected animals with increasing dose (i.e. an increase in incidence), given a specific severity level (e.g. mild lesion). The rate of that increase with dose depends on the variability among the animals, as well as experimental errors (e.g. dosing errors, cage effects). This concept is illustrated in Fig. 3.2. Because the observed quantal responses are conceptualized as being “determined” by the cut-point applied to an underlying continuous response, effects under this interpretation are referred to as deterministic quantal end-points.

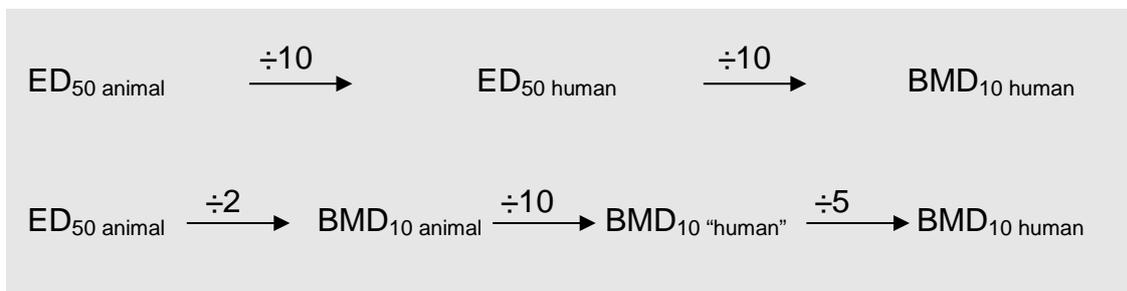
Because the shape of the quantal dose–response curve in this case reflects animal intraspecies variability and experimental variation, this particular aspect of the data is less relevant for humans. Therefore, (human) intraspecies variability could be fully covered by taking the median effective dose ( $ED_{50}$ ) as a PoD and applying an assumed factor (or distribution) for human variability, based on human data (see section 4.5). This could lead to the same result as a procedure that starts from an  $ED_{10}$  ( $BMD_{10}$ ) and applies a factor that arises from assuming that the human  $ED_{50}/ED_{10}$  ratio is  $x$ -fold greater than that of the test animal (see Fig. 3.3 for an illustration). Whatever procedure is used, the extrapolated dose still relates to the severity level  $M$  as defined by the underlying quantal data (e.g. mild lesion).

Therefore, in the deterministic interpretation of quantal dose–response data, the magnitude of effect  $M$  may be a qualitative descriptor of the effect magnitude (e.g. mild or more severe, moderate or more severe, etc.). An example of this in the context of specifying  $HD_M^I$  is:

- ✚  $HD_{mild}^{05}$  (critical effect = liver lesions): the human dose at which 5% of the population shows mild or more severe liver lesions.



**Fig. 3.2: Deterministic quantal end-points: quantal responses reflecting incidences of a continuous response above and below a fixed cut-point.** When this cut-point is imagined to relate to, say, a mild lesion, the associated median effective dose ( $ED_{50}$ ) in the lower panel relates to the dose at which 50% of the animals are observed to show (at least) mild lesions. The quantal dose–response curve around the  $ED_{50}$  in the lower panel reflects the fraction of (hypothetical) observations exceeding  $M^*$  in the scatter distributions in the upper panel.

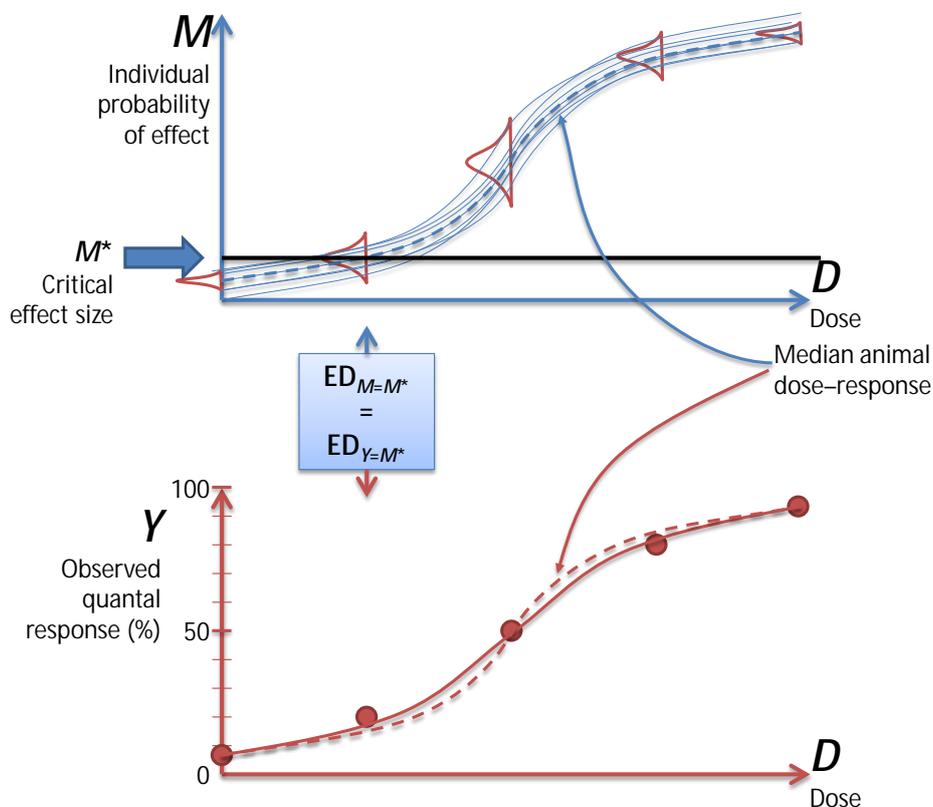


**Fig. 3.3: Example of an extrapolation from an animal  $ED_{50}$  or animal  $BMD_{10}$  to a human  $BMD_{10}$  in the case of a quantal end-point.** When the ratio  $ED_{50}/BMD_{10}$  in the animal is assumed to equal 2, an inflating factor of 5 for larger human variability is equivalent to applying an intraspecies factor of 10 directly to the animal  $ED_{50}$ . The  $BMD_{10}$  "human" reflects the hypothetical  $BMD_{10}$  in a human population having the same variability as the test species in the particular experiment.

Thus, 95% of the population would be protected from mild or more severe liver lesions. Holding severity  $M$  constant, lower doses (lower values of HD) will result in a lower fraction  $I$  of the population experiencing an effect of severity  $M$  or greater. Holding  $I$  constant, lower doses will result in a lower severity of effect, such as a “minimal” rather than “mild” lesion, in this fraction of the population.

For quantal data relating to cancer (and possibly other end-points, such as malformations), another interpretation of the observed incidences is possible. Here, the incidences could also be interpreted as reflecting the individual probability of developing cancer at the tested dose.

If so, this individual probability can be used as a measure of severity ( $M$ ) at the level of an individual. For instance, in carcinogenicity studies (which often use homogenous strains), it might be assumed that the animals are close replicates and that each individual animal has approximately the same cancer probability when receiving the same dose. Specifically, when the animals in a dose group all have an individual cancer probability of 0.20, then the expected incidence would be 20%. The impact of this interpretation is that the steepness of the dose–response curve can now be considered as a characteristic of the chemical, rather than of the studied animal population (as it is in incidences of histopathological effects). Hence, a  $BMD_{10}$  (for a benchmark response of 10% extra risk) could be considered as an estimate of the dose at which the typical human being would be subject to a 10% extra cancer risk (apart from interspecies extrapolation). This concept is illustrated in Fig. 3.4, as a contrast to Fig. 3.2. Because the observed quantal response is conceptualized as random, or “stochastic”, with the probability of effect viewed as the underlying continuous response, effects under this interpretation are referred to as stochastic quantal end-points.



**Fig. 3.4: Stochastic quantal end-points: quantal responses reflecting individual probability of effect.** The dashed dose–response curve in the lower panel is the median animal curve in the upper panel, but the “observed” uninterrupted curve in the lower panel is shallower due to the variability among animals (see distributions in the upper panel).

Therefore, in the stochastic interpretation of quantal dose–response data, the magnitude of effect  $M$  may be a quantitative level reflecting an individual probability of effect (e.g. 10% extra risk). Because the shape of the quantal dose–response curve in this case reflects a measure of response at the level of the individual, (human) intraspecies variability needs to be accounted for in order to address differences in sensitivity across individuals (i.e. differences in the dose that would elicit the same individual probability of effect). Therefore, an example of this in the context of specifying  $HD_M^I$  is:

- ✚  $HD_{05}^{01}$  (critical effect = extra risk of lung tumours): the human dose at which there is an individual extra risk of lung tumours of 5% (or more) in 1% of the population.

Thus, 99% of the population would be protected from a 5% or greater individual extra risk of lung tumours. Holding severity  $M$  constant, lower doses (lower values of HD) will result in a lower fraction  $I$  of the population experiencing an effect of severity  $M$  or greater. Holding  $I$  constant, lower doses will result in a lower severity, such as a 1% rather than a 5% extra risk of lung tumours, in this fraction of the population.

Although the definition of the  $HD_M^I$  for the case of stochastic quantal end-points is fully in line with the overall framework, risk managers may be more interested in the overall rate of cases in the whole population. That value can be directly derived by integrating the  $HD_M^I$  over all possible values of  $M$  (see Slob et al., 2014).

### **3.2.3 Principle 3: The concept of an “effect metric” for $M$ forms the basis of “equipotency” and differences in “sensitivity”**

As already noted, hazard characterization is the process of deriving a PoD from dose–response data and extrapolating that dose to an equipotent dose in the target population (and to target exposure conditions). Two doses are equipotent when they result in the same effect size  $M$  in two individuals, which might be two typical individuals, each representing a different species, or two individuals of the same species that differ in sensitivity. Therefore, it must be clear what is meant by the “same effect size” – i.e. how to define the effect metric that forms the basis of “equipotency” and differences in “sensitivity”.

A key question for defining the effect metric is how to address differences in background response. For instance, when extrapolating a PoD from an experimental animal study to an equipotent dose for the typical human being, the background response in the control group of the animal study needs to be taken into account. This is easy when the critical end-point is continuous: risk assessors express the magnitude of an effect in terms of a per cent change (e.g. 5% decrease in red blood cell count, 10% increase in relative liver weight). It is thus assumed that a per cent change in a continuous end-point reflects an equivalent effect size in the study and target populations, even when their background responses differ. After appropriate interspecies adjustment, a dose is obtained at which the typical human being is assumed to be subject to that same effect size (per cent change).

For quantal end-points, defining the effect metric depends on the interpretation of quantal data (see section 3.2.2 above). For a deterministic quantal end-point, when the dose–response is interpreted as resulting from experimental variation (including experimental errors, such as dosing errors, and remaining genetic differences among animals), the appropriate PoD would be an  $ED_{50}$ . The  $ED_{50}$  represents the dose at which the median experimental animal is subject to the effect considered. In this definition, the background response does not occur<sup>10</sup> and does not interfere when extrapolating the  $ED_{50}$  in the test species to the  $ED_{50}$  in humans (unless the observed background response is close to or higher than 50%). In simpler words, if the underlying gradual effect is converted into a “yes” or “no” (equal to or greater than/less than) response, then the  $ED_{50}$  corresponds to the dose required to elicit a “yes” response in the median animal.

However, for a stochastic quantal end-point, when the quantal dose–response relationship is interpreted as the dose-dependent individual risk, an equivalent measure of risk needs to be defined between the test animals and the human population. As the two populations may

---

<sup>10</sup> Histopathologists tend to define the category “no lesion” based on the observed tissues in the controls, resulting in low observed background incidences in most cases. Yet high background responses may occur, for example, when there are sex differences.

differ in background risk at zero exposure, the question is how to correct for that in such a way that the resulting change in risk may be interpreted as an equivalent change. Various measures exist, such as additional and relative risk, but none of them seems to meet the requirement of representing an equivalent change in risk when the background response differs substantially among the two populations. For instance, suppose the animals show an increase in incidence from 10% to 20%, whereas the background incidence in humans is 1%. Then, using the additional risk metric, the equipotent human dose would relate to an incidence of 11%, but using the relative risk metric, only 2%. This example illustrates that the prediction of risks is highly dependent on the risk metric used.

The use of either measure appears to be based on discipline-associated habits (toxicologists versus epidemiologists) or risk management paradigms (precautionary principle), whereas there are few scientific arguments to support the preference of one measure over the other. Yet the public health impacts of these different measures of risk may be substantial and can be of great importance for choices among policy options.

After the animal PoD is extrapolated to an equipotent human dose for the typical individual, this dose needs to be extrapolated to an equipotent dose for the sensitive human being. In this context, the term “sensitive individual” is used to indicate that the same effect size  $M$ , using the same effect metric, would be evoked at a lower dose in these individuals than in less sensitive individuals. In keeping with the  $HD_M^1$  concept, the terms “typical” and “sensitive” need to be made more explicit. Therefore, from this point forward, the term “typical” will refer specifically to the “median” (50th percentile in the human population), whereas the term “sensitive” will relate to an incidence  $I$  less than 50% in the population (e.g. an incidence of 5% refers to the 5th percentile of the population).

It should be noted here that equipotent doses in different subjects may not be equally adverse. For instance, a 5% decrease in red blood cell count will have more impact in persons with anaemia than in healthy individuals. The individuals with anaemia might be denoted as “vulnerable”. Vulnerable individuals might be more, less or equally sensitive to the chemical. For instance, vulnerable individuals with anaemia might be equally sensitive – that is, they might show the same per cent decrease in red blood cell count at the same dose as an individual with a typical background value, even if the health impact caused by the same 5% change is greater in individuals with anaemia than in individuals with “normal” background values.

### **3.2.4 Principle 4: Making inferences from a point of departure involves making adjustments while accounting for uncertainty and variability**

#### *3.2.4.1 Uncertainty versus variability*

Although both uncertainty and variability are often described by statistical distributions, they are fundamentally different concepts. Uncertainty relates to “lack of knowledge” that, in theory, could be reduced by better data, whereas variability relates to an existing aspect of the real world that is outside our control. For instance, the factor for intraspecies variability is adjusting for the differential sensitivity within the human population. At the same time, we are uncertain about the degree to which people might differ in their equipotent doses – that is, the individual doses needed to elicit a specific response with a specific degree of severity (magnitude). The differential sensitivity is non-reducible variability, but our uncertainty about that variability might be reduced if we had the appropriate data.

#### *3.2.4.2 Adjustment and uncertainty*

In most scientific work, uncertainty relates to a parameter that is assumed to have a single unknown value. This value is estimated based on data, where better data will result in a better estimate (i.e. with less uncertainty). This principle applies to the BMD, the value of which is estimated by the dose–response data, where better data result in less uncertainty in that estimate (as reflected by a smaller BMD confidence interval). Most of the other aspects of hazard characterization, however, involve extrapolation from one situation to another. Here, the uncertainty in that extrapolation is informed by variability among chemicals as observed in historical data. In other words, in the extrapolation aspects, variability among chemicals translates into uncertainty for the specific chemical to be assessed.

For instance, one of the steps may be that the PoD for a subchronic exposure needs to be adjusted to an equipotent dose for a chronic exposure. From historical data, it is known that equipotent doses between subchronic and chronic studies differ by a factor that varies among chemicals (see section 4.3). This implies that the appropriate assessment factor depends on the specific chemical. Therefore, if one wishes to provide for cases where the chemical under consideration has a larger subchronic to chronic ratio than is typical, the factor to be applied (in the absence of chemical-specific data) needs to be high enough to cover most chemicals. Thus, in a deterministic hazard characterization, a data-based assessment factor for this aspect may be conceived as the assessment factor for a chemical in the high-end tail of the distribution. In contrast, in a probabilistic hazard characterization, the whole distribution is used in the calculations. The width of this distribution will not be reduced by collecting more data (e.g. additional chemicals with subchronic to chronic PoD ratios available), as the distribution reflects true variability among chemicals.<sup>11</sup> In this situation, the only way to reduce uncertainty is to replace this distribution with one based on chemical-specific data (see section 4.9).

### 3.3 Applying the framework

#### **3.3.1 Quantifying uncertainties individually and in combination**

##### *3.3.1.1 Quantifying uncertainties individually*

Common aspects of hazard characterization, each of which has uncertainty, were listed in section 3.1. Hazard characterization typically begins with the identification of a PoD. Which of the other aspects are involved in any specific case depends on the situation. Some aspects are frequently involved (e.g. interspecies and intraspecies differences), others less frequently (e.g. exposure duration, route-to-route extrapolation); still others are “emerging” (e.g. in vitro to in vivo; read-across).

In current practice, the uncertainty in these aspects is typically taken into account by using assessment factors, often default factors. These default factors are generally considered to be conservative values. For instance, a factor of 10 for interspecies differences does not represent the assumption that humans are typically a factor of 10 more sensitive than test animals, but rather that they are not more than 10 times more sensitive in the majority of cases. This implies that humans are expected to be less than 10 times more sensitive in most cases; put another way, the factor of 10 is considered to be an upper-bound estimate of the potential interspecies difference (for the specific chemical considered).

---

<sup>11</sup> Including more chemicals will, of course, lead to a better specified distribution and a more precise estimate of the across-chemical variability.

In evaluating uncertainties, however, not only an upper-bound (percentile) but also a lower-bound (percentile) estimate of the underlying uncertainty distribution is needed. In principle, the lower-bound estimates for interspecies differences could include values smaller than unity, if it were assumed that test animals could actually be more sensitive than humans. Wherever possible, the uncertainty ranges of assessment factors should be informed by experimental (historical) data.

As section 4 shows, data or methods are available to estimate or inform the uncertainty distributions of many of the aspects of hazard characterization (including the PoD). Once uncertainties in the individual aspects are estimated, there are a number of ways in which they can be combined, as discussed next.

### *3.3.1.2 Combining uncertainties*

A number of approaches to combining uncertainties are possible. As a primitive approach, one may multiply all conservative and non-conservative bounds of the uncertainty ranges for each aspect, resulting in an uncertainty range for the final outcome of the hazard characterization (the lower bound of which could be the RfD, such as an acceptable daily intake [ADI] or tolerable daily intake [TDI]). This final range, however, would be overly wide in the sense that the final uncertainty is overestimated (i.e. the “coverage” of the final result will be greater than the “coverage” of each component). A more accurate way of combining the uncertainties over all aspects involved is by reflecting the uncertainty for each aspect by an uncertainty distribution and then combining these distributions using probabilistic methods (e.g. Monte Carlo simulation). The resulting distribution will better reflect the degree of uncertainty in the final outcome of the hazard characterization. In addition, from this distribution, a value can be selected associated with the coverage level that is considered adequate. In this way, the level of conservatism (in the sense of coverage) remains under control.

In summary, this document presents three basic options for how to combine uncertainties:

1. *Non-probabilistic analysis* – This is where the individual lower and upper bounds for each hazard characterization aspect are combined by multiplication. The lower and upper bounds are data based – that is, chosen percentiles from uncertainty distributions, such as discussed in section 4. In this approach, the resulting bounds for the combined uncertainty cannot be interpreted in terms of coverage, other than to say that the overall coverage is greater than the coverage of the individual aspects of hazard characterization.
2. *Approximate probabilistic analysis* – This is where uncertainty distributions are combined probabilistically. Uncertainties in PoDs and in assessment factors are all assumed to be independently lognormally distributed, so the calculations may be performed without Monte Carlo simulations, making the calculations easier (see Annex 1).
3. *Full probabilistic analysis* – Although the overall framework for analysis is the same as for the approximate probabilistic approach, a full probabilistic analysis is more flexible, as summarized below (see also Annex 1 for an overview).

The first, non-probabilistic approach is analogous to the traditional approach of multiplying factors considered to be conservative, but includes the additional step of specifying opposite (non-conservative) bounds. Further, the “data-based” factors used may differ from the traditional default factors and be based on historical data, as reviewed in section 4. This method of multiplying bounds is simple, but at the expense of an unknown probability that

the two final bounds include the target value (see also Annex 1). As a second disadvantage, the method tends to be highly conservative in the sense of a relatively large distance from the lower bound to the upper bound, in particular with increasing number of aspects.

The second option involves an approximate probabilistic analysis, which can easily be performed in a Microsoft Excel spreadsheet, by assuming that all uncertainties can be reflected by statistically independent lognormal distributions (see Annex 2). A prototype software tool (“APROBA”) has been developed in conjunction with this document, which is described in the next section (for user instructions, see Annex 2).

Example calculations in the case where only interspecies and intraspecies aspects are needed using both non-probabilistic and approximate probabilistic approaches are shown in Fig. 3.5.

The third option is to conduct a case-specific full probabilistic hazard characterization (see Annex 1). The main differences between this option and the approximate probabilistic analysis are as follows:

- the uncertainty distribution for the BMD can be evaluated for the specific data set and models used (e.g. by bootstrapping; Moerbeek, Piersma & Slob, 2004), rather than assuming it to be lognormal;
- the uncertainty in intraspecies variability more closely corresponds to the assumptions made in the analysis of the underlying data (Hattis & Lynch, 2007);
- uncertainties other than those in the usual aspects can be included (e.g. in PBTK model parameters);
- correlations among uncertainties in different aspects of hazard characterization can be incorporated;
- uncertainties at different levels of  $M$  can be evaluated; and
- uncertainties in the exposure assessment parameters can be included at the same time, resulting in an integrated probabilistic risk assessment, or IPRA (Van der Voet & Slob, 2007).

The full probabilistic approach is the most accurate and most flexible. Currently, however, no general user-friendly software package is available, and the assessor needs to compose his or her own software, which in general includes Monte Carlo simulations. For that reason, the approximate probabilistic analysis will in many cases be a helpful alternative.

### **3.3.2 Spreadsheet tool “APROBA” for facilitating approximate probabilistic analysis**

As part of this Harmonization Project, a spreadsheet tool called “APROBA” (for “Approximate PROBabilistic Analysis”)<sup>12</sup> has been developed in Microsoft Excel to facilitate such analyses and to ensure wide accessibility (see Annex 2 for a user guide).

---

<sup>12</sup> Available online on the WHO/IPCS Harmonization Project publications webpage at: <http://www.who.int/ipcs/publications/methods/harmonization/en/>

A

$$HD_M^I = \frac{BMD_M}{AF_{Inter-BS} \cdot AF_{Inter-TK/TD} \cdot AF_{Intra-I}}$$

B

Non-probabilistic analysis		
Aspect	P05 (lower confidence limit)	P95 (upper confidence limit)
BMD <sub>M</sub>	100	900
AF <sub>Inter-BS</sub>	3.68	5.49
AF <sub>Inter-TK/TD</sub>	0.333	3
AF <sub>Intra-I=1%</sub>	2.24	41.9
	Lower confidence limit	Upper confidence limit
HD <sub>M</sub> <sup>0.01 (I=1%)</sup>	$\frac{100}{(5.49 \times 3 \times 41.9)} = 0.14$	$\frac{900}{(3.68 \times 0.333 \times 2.24)} = 328$

C

Approximate probabilistic analysis		
Aspect	P50 (median)	P95/P50 (uncertainty)
BMD <sub>M</sub>	300	3
AF <sub>Inter-BS</sub>	4.50	1.22
AF <sub>Inter-TK/TD</sub>	1	3
AF <sub>Intra-I=1%</sub>	$(2.24 \times 41.9)^{1/2} = 9.69$	$(41.9 \div 2.24)^{1/2} = 4.32$
HD <sub>M</sub> <sup>0.01 (I=1%)</sup>	$\frac{300}{(4.50 \times 1 \times 9.69)} = 6.9$	$\frac{[(\log 3)^2 + (\log 1.22)^2 + (\log 3)^2 + (\log 4.32)^2]^{1/2}}{10} = 8.5$
	P05 (lower confidence limit)	P95 (upper confidence limit)
HD <sub>M</sub> <sup>0.01 (I=1%)</sup>	$6.9 \div 8.5 = 0.81$	$6.9 \times 8.5 = 59$

**Fig. 3.5: Example calculations of both non-probabilistic and approximate probabilistic approaches in the case where only interspecies and intraspecies aspects are needed.** A. Basic equation for HD<sub>M</sub><sup>I</sup> in the case where the PoD is a BMD and only interspecies and intraspecies aspects are needed. B. Non-probabilistic calculation of uncertainty. C. The approximate probabilistic calculation of uncertainty. Note: Uncertainty distributions are based on the preliminary distributions described in section 4. The approximate probabilistic analysis assumes that AF<sub>Intra-I</sub> is lognormally distributed, with the same 5th percentile (P05) and 95th percentile (P95) values as for the original distribution. Thus, it does not use the 50th percentile (P50) values and P95/P50 from the original distribution, but rather uses the P50 and P95/P50 calculated from the lognormal approximation. Note that the values used for AF<sub>Intra-I</sub> in this table can also be read from table 4.5.

This tool allows evaluation of the uncertainty for many typical hazard characterizations, resulting in approximate values for coverage and degree of uncertainty in the derived estimate of the HD<sub>M</sub><sup>I</sup>. Fig. 3.6 shows a screenshot of the tool, applied to the same example calculations as shown in Fig. 3.5.

INPUTS RELATED TO STUDY, END-POINT AND PROTECTION GOALS			
DESCRIPTION	INPUTS	COMMON VALUE(S)	NOTE
End-point	Example calculation	Case-specific	
Data type	Continuous	Case-specific	
Data route	Oral	Case-specific	
Study type	Chronic	Case-specific	
Test species	Rat	Case-specific	
Body weight test species	0.400	0.4	a
Human median body	60.0	60	
Target BMR (= AF, user input for BMDs only)	5%	5%	b
Population incidence goal	1%	5%, 1%, 0.1%, 0.01%	
Probabilistic coverage	95%	95%	
PoD type	BMDL	Case-specific	
PoD value	100	Case-specific	
BMDU (User input for BMDL)	300	Case-specific	c
PoD units	mg/kg body weight per day	mg/kg body weight per day	
Deterministic overall AF	100	Case-specific	
Deterministic RID	1.0	Calculated	
Exposure estimate	1.0	User supplied	

INPUTS RELATED TO ADJUSTMENT, VARIABILITY AND UNCERTAINTY			
HAZARD	INPUTS	PROVISIONAL VALUE(S)	NOTE
PoD	LCL	100	Calculated from inputs
(Modelled BMD uncertainty)	UCL	300	Calculated from inputs
NOAEL to BMD	LCL	1.00	
(NOAEL or LOAEL only)	UCL	1.00	
Interspecies scaling (Allometric for oral)	LCL	3.68	
	UCL	5.49	
Interspecies TK/TD (Remaining TK & TD)	LCL	0.333	
	UCL	3.00	
Duration extrapolation	LCL	1.00	
	UCL	1.00	
Intraspecies	LCL	2.24	
	UCL	41.88	
Other aspect #1 (Description here)	LCL	1.00	
	UCL	1.00	
Other aspect #2 (Description here)	LCL	1.00	
	UCL	1.00	
Other aspect #3 (Description here)	LCL	1.00	
	UCL	1.00	

INTERMEDIATE CALCULATIONS FOR UNCERTAINTY ANALYSES			
ASPECT		[log(P95/P50)]	% contribution to overall
PoD	P50	3.00	26%
	P95/P50	3.00	0.228
NOAEL to BMD	P50	1.00	---
	P95/P50	1.00	0
Interspecies scaling	P50	4.50	1%
	P95/P50	1.22	0.00758
Interspecies TK/TD	P50	1.00	26%
	P95/P50	3.00	0.228
Duration extrapolation	P50	1.00	---
	P95/P50	1.00	0
Intraspecies	P50	3.68	47%
	P95/P50	4.32	0.404
Other aspect #1 (Description here)	P50	1.00	---
	P95/P50	1.00	0
Other aspect #2 (Description here)	P50	1.00	---
	P95/P50	1.00	0
Other aspect #3 (Description here)	P50	1.00	---
	P95/P50	1.00	0
(HDI <sub>m</sub> ) <sup>1</sup>		Non-Prob. 6.89, Approx. Prob. 6.89	Greatest uncertainty
	UCL/P50	47.6	Intraspecies

NON-PROBABILISTIC ANALYSIS OUTPUTS <sup>1A</sup>			
(HDI <sub>m</sub> ) <sup>1</sup>	LCL	0.145	day
	UCL	328	mg/kg body weight per
Fold Range of Uncertainty		2261	
LCL of HDI <sub>m</sub> <sup>1</sup>			100%

\*Based on approximate probabilistic analysis, below.

APPROXIMATE PROBABILISTIC ANALYSIS OUTPUTS			
Standard Confidence			
(HDI <sub>m</sub> ) <sup>1</sup>	LCL (P05)	0.807	day
	UCL (P95)	58.8	mg/kg body weight per
Degree of Uncertainty (if old Range)			73
Estimated "Coverage" of Deterministic			93%
Probabilistic RID			
0.81	= Approximate probabilistic HDI <sub>m</sub> <sup>1</sup> at specified % confidence		
	= Estimate of dose (mg/kg body weight per day) at which, with		
	95% confidence		
	1% of the population will have	Example calculation	
	of	>	5%

GENERAL APPROACH			
<b>Non-probabilistic analysis</b> multiplies together conservative or non-conservative confidence limits (P05 or P95) for each uncertainty.			
<b>Approximate probabilistic (Approx. Prob.) analysis</b> combines uncertainties probabilistically assuming independent lognormal distributions.			
Defines Lower Confidence Limit (LCL) = P05, Upper Confidence Limit (UCL) = P95. Given P50 and P95/P50, assumes P05 = P50/(P95/P50). Given P05 and P95, assumes P50 = sqrt(P05*P95).			
USER			
[User can enter any notes here]			

**NOTES:**

- a - Automatically adjusts for mice, rats, dogs, and rabbits.
- b - For NOAEL/LOAEL, is 5% if continuous and 10% if quantal-stochastic and 50% if quantal - User input is ignored if NOAEL/LOAEL or quantal-deterministic. Otherwise user inputs BMR used for
- c - For NOAEL/LOAEL, PoD is fixed.
- d - For BMD, assumes LCL=BMDL, UCL=BMDU.
- e - Uncertainty in NOAELs as surrogate for BMD.
- f - For deterministic quantal effects, also includes adjustment from NOAEL to ED<sub>50</sub>.
- g - Allometric scaling for oral dosing using user input body weights. User must supply for inhalation or dermal.
- h - Accounts for case-specific deviation from the general interspecies scaling.
- i - Depends on population incidence protection goal.
- j - For user defined value, specify LCL and UCL on Log(CSD<sub>01</sub>), then calculate the Intraspecies LCL = 10\*(NORMSINV(1-C13))\*Log(CSD<sub>01</sub>), where cell C13 contains the population incidence protection
- k - Can add other extrapolation aspects (e.g., LOAEL-to-NOAEL), as long as P05 and P95 are
- l - Non-probabilistic LCL = LCL on PoD / Product of UCLs of Each Aspect.
- m - Non-probabilistic UCL = UCL on PoD / Product of LCLs of Each Aspect.

Fig. 3.6: Screenshot of APROBA tool<sup>13</sup> for performing an approximate probabilistic analysis, provided together with this guidance. APROBA conducts calculations using both non-probabilistic and approximate probabilistic approaches. In the example here, the PoD is a BMD, and only interspecies and intraspecies aspects are needed.

The current version of APROBA includes default uncertainty distributions (which can also be changed) for the aspects PoD, interspecies and intraspecies extrapolation, and exposure duration extrapolation. These default distributions are taken from section 4 (Table 4.6). The spreadsheet further includes some fields called “other aspects”, where the user can specify the uncertainty for any other aspect for which no distribution has been proposed so far. When the user is able to quantify the uncertainty of any aspect in terms of two values – for example, a 5th percentile (P05) and a 95th percentile (P95) value – this translates into a lognormal distribution that can be included in the probabilistic analysis. These two values may be estimated based on any data available, in some cases based on expert judgement (see section 3.4 for a brief discussion of these “other” uncertainties).

Stepwise application of the APROBA tool to a real-life example is also demonstrated in the case-study on deoxynivalenol included in Annex 5 of this guidance.

<sup>13</sup> This screenshot was generated with APROBA version 1.0. Note that in the future the APROBA tool might be further developed without changing this guidance document every time. As a consequence, slight discrepancies between the screenshots in this document and the most recent version of APROBA might occur.

### 3.4 Evaluation of other uncertainties

Table 3.1 listed a number of aspects for which the associated uncertainties could, at least in principle, be quantified based on data, and most of them with the approach of evaluating PoD ratios from historical data. For some of these uncertainties, section 4 discusses data that could be used as the basis for preliminary, generic defaults. No preliminary defaults have been proposed for the other uncertainties, but they can be quantified by the risk assessor whenever needed. Section 4.8 provides some discussion on how those uncertainties could be quantified. In this section, some specific “other” uncertainties not mentioned in Table 3.1 are reviewed.

#### **3.4.1 Uncertainties regarding distributions used in the probabilistic approach**

Obviously, the results from a probabilistic analysis are valid to the extent that the assumed uncertainty distributions adequately describe the uncertainty for that aspect. The generic distributions used to represent some sources of uncertainty (introduced above) are based on data that vary in quantity, quality, relevance, comprehensiveness, etc. (see section 4). This means that there is uncertainty about how well the generic distributions represent the uncertainties they address, and that uncertainty may, for some aspects, be larger than for others. This additional uncertainty may arise from limitations in the suitability of the form of distribution (generally assumed to be lognormal) used to represent each uncertainty and from uncertainties associated with the parameters used to define those distributions (i.e. upper and lower bounds or central tendency and spread). These types of uncertainty could be regarded as “secondary” in the sense of representing uncertainty about the distributions used to quantify uncertainty. The fact that secondary uncertainties exist implies that the outputs of the probabilistic analysis are themselves uncertain. For example, the calculated coverage would have been lower or higher had the quantified uncertainties been assumed to be larger or smaller or had additional sources of uncertainty been included or not. Therefore, even though a probabilistic analysis provides a better characterization of uncertainties compared with the traditional deterministic approach, the results should not be considered exact.

In principle, it is possible to quantify secondary uncertainty and incorporate it into the probabilistic analysis. The sampling uncertainty in the parameters of the distribution (geometric mean [GM] and geometric standard deviation [GSD]; see section 4) that arises from estimating a distribution from data for a limited number of chemicals can be quantified relatively easily, especially in the case of lognormal distributions, for which suitable equations exist (e.g. Vose, 2008). Quantifying other secondary uncertainties, such as uncertainty about the relevance or representativeness (e.g. in chemical space) of the available data, is more difficult and is likely to require expert judgement. It would be good to include quantification of the secondary uncertainties in any future work aiming at estimating distributions from data, including refinement of already proposed generic distributions or development of new distributions for aspects not yet considered (see recommendations for research in section 7). A practical approach to address this type of uncertainty for the time being can be found in the case-study provided in Annex 5.

#### **3.4.2 Uncertainties regarding the assumption of independence**

When combining different uncertainties that have been quantified probabilistically in the approximate probabilistic approach, it is assumed that they are all independent of one another (see section 3.3 above). In other words, it is assumed that new information about the true value for one component of the assessment (e.g. the interspecies differences) would not alter the uncertainty of the other components (e.g. intraspecies variability). Any uncertainty about

this assumption would be an additional source of uncertainty affecting the results of the probabilistic analysis. However, the assumption of independence is considered reasonable for the elements quantified probabilistically in the present document and a negligible source of uncertainty (see section 4.10).

### **3.4.3 Uncertainties relating to qualitative aspects of hazard assessment**

This document focuses mainly on uncertainties relating to entities that assume quantitative values, such as PoDs and assessment factors. However, uncertainties may also be involved in non-quantitative aspects of hazard assessment, such as when assessing whether a given end-point is adverse and relevant for humans and hence whether it should be considered as relevant for risk assessment. Clearly, these are chemical-specific uncertainties. In the case of relevance to humans, mode of action considerations come into play. Depending on the amount of information available on the mode of action or on species-dependent mechanisms, the statement that the (potentially critical) end-point is relevant for humans can be given a specific probability (expressing the primary uncertainty). Note that an uncertainty range around that probability would be analogous to the secondary uncertainty discussed in section 3.4.1 above.

The specified probability would translate into an additional uncertainty measure for a derived health-based guidance value, independent from the uncertainties about the  $HD_M^I$  considered above. This uncertainty measure would indicate if the derived health-based guidance value is at all relevant for consideration.

### **3.4.4 Uncertainties that are difficult to quantify**

Some other types of uncertainty not mentioned so far, but which are usually difficult to quantify, may be involved in a hazard characterization, such as:

- uncertainties regarding responses that are not captured by the BMD confidence interval – for instance, litter effects when data to take litter effects into account are not reported, or other deficiencies and limitations in the reporting of data;
- limitations in the scientific quality or reporting of a study (i.e. regarding the experimental protocol); and
- uncertainty in the PoD related to interstudy variation (e.g. variation over several 90-day studies with the same chemical in the same species for the same effect or in different species).

In all such situations, one might consider postulating an uncertainty distribution based on expert judgement. In this way, the uncertainty can be included in the probabilistic assessment in the usual way (the APROBA tool includes options for this; see section 3.3.2 and Annex 2). When distributions are derived by expert judgement, it is essential to make this explicit, in order to distinguish them from distributions estimated from data (USEPA, 1997) and to document the evidence and reasoning on which the distributions are based. Expert judgements are subject to several types of cognitive bias; thus, where the resulting distributions are critical to the assessment, it is advisable to derive them using formal methods of expert elicitation, which are designed to reduce bias (e.g. EFSA, 2014). Although arrived at subjectively, the results will still be of great value to a risk manager (USEPA, 2000).

The last type of uncertainty in the list above (interstudy variation) needs some separate discussion. It is a complex issue, and it should be noted that interstudy variability can result

from many different sources of variability, such as differences in administration of the dose, in experimental conditions (either in the same laboratory or in different laboratories) or in strains or age of the animals.

Where only a single study exists for the critical end-point, the interstudy uncertainty is not apparent and might be mistaken as non-existing. In reality, however, the interstudy uncertainty in the PoD is largest in the case of only one study and decreases with more studies, although it is unknown to what extent.

The (possibly hidden) presence of interstudy variability implies that the PoD chosen from a single study might have been less or more conservative than those derived from other studies (had they been available), but without knowing to what extent. The probabilistic approach in this document was developed to avoid the latter and to make the level of conservatism visible. However, quantifying the uncertainty in PoDs due to variation among studies is difficult, and a satisfactory general solution appears to be lacking. One of the difficulties is that the number of available studies varies from case to case; the quality of the studies plays a role as well. Lack of information on species specificity adds to the problem. In cases where there are multiple studies for the same end-point, a study with a relatively high PoD might be preferred, due, for example, to concerns about the design or conduct of the studies with lower PoDs. However, unless the studies that would give lower PoDs are considered so flawed as to provide no information at all, omitting them from the hazard characterization will tend to result in a relatively non-conservative PoD.

In any specific assessment, there may still be other uncertainties not mentioned in this document that are difficult to quantify as well. In principle, all uncertainties identified should be taken into account somehow, preferably in a quantitative way. Even rough quantification of uncertainty is better than ignoring the uncertainty, which amounts to assuming that the uncertainty is absent. If quantifying these uncertainties is deemed totally inappropriate, then accounting for them qualitatively when interpreting the  $HD_M^I$  might be considered. In practice, this means describing the potential impact in words. A disadvantage of this is that words are interpreted differently by different people, and so the risk manager may overinterpret or underinterpret the described impact. Therefore, a better option is to perform a sensitivity analysis, by re-evaluating the overall uncertainty of the current hazard characterization based on different quantitative estimates of the uncertainty of the “difficult-to-quantify” aspect. This approach is illustrated in the case-study on deoxynivalenol (Annex 5).

USEPA (1997) guidance for Monte Carlo analysis states that “There are limits to the assessor’s ability to account for and characterize all sources of uncertainty. The analyst should identify areas of uncertainty and include them in the analysis, either quantitatively or qualitatively.” In the longer term, more research is needed to develop methods to evaluate the more difficult uncertainties in a quantitative way to the extent possible. Section 7 summarizes a number of research needs that appear most urgent at this point.

## 4. IMPLEMENTATION OF THE APPROACH: PRELIMINARY UNCERTAINTY DISTRIBUTIONS

To implement the framework, we need uncertainty distributions for all aspects. In many cases, there is no chemical-specific information on the assessment factor to be used, and the uncertainty in that aspect is generic. In those situations, we can use only generic uncertainty distributions, analogous to the generic assessment factors in the deterministic hazard characterization. For some of the aspects, historical data are available that can be used for informing generic uncertainty distributions. Sections 4.1–4.6 review proposed uncertainty distributions for a number of aspects, which may be regarded as preliminary defaults in the absence of a more extensive systematic review. Section 4.7 summarizes these distributions (Table 4.6). Section 4.8 mentions some of the hazard characterization aspects that might be quantified, but for which no attempt has been made so far to translate historical data into a preliminary uncertainty distribution. Cases where chemical-specific information may be used for deriving a CSAF distribution are discussed in section 4.9, including the consequences for the uncertainty distributions of other hazard characterization aspects. Section 4.10 discusses secondary uncertainties – that is, the issue that the uncertainty distributions themselves are uncertain, depending on the quality of the data that informed them.

### 4.1 Using historical data: general approach

The overall approach for defining an uncertainty distribution for a given aspect is to search for relevant data in the literature, usually from studies that reanalysed or reviewed a particular set of historical data. The general idea of using historical data in a generic way for informing the uncertainties involved in each aspect applies to most aspects. For example, the uncertainty in the subchronic PoD as a surrogate for the chronic PoD can be informed by a histogram of observed PoD ratios between subchronic and chronic studies for the same chemicals.

Uncertainty distributions derived in this way are called “generic uncertainty distributions” in this document. The general assumption is made that for a given aspect, the historical data (e.g. PoD ratios) used to inform these uncertainty distributions follow a lognormal distribution. Theoretically, it might be argued that the form of the distribution is uncertain itself and should be included in the uncertainty evaluation. However, distributions of NOAELs or BMD(L)s are found to be consistent with lognormality (see the review studies mentioned below – e.g. Bokkers & Slob, 2007), and so are observed ratios of PoDs (note that theoretically the ratio of two lognormal distributions is again lognormal). Therefore, this assumption appears to be generally useful, and the impact of possible deviations is likely to be small outside of the extreme tails. Moreover, the uncertainty in the distribution shape is not likely to be as important as the uncertainties in the GM and GSD estimates based on available data, which are often limited.

Given the assumption of lognormal ratios, the GM can be regarded as an estimate of the median of the distribution (P50). Next to the median (P50 = GM), the distributions will be characterized by the ratio between the P95 and P50 values as the measure of its spread (note that  $P95/P50 = GSD^{z\text{-score at } 0.95}$ ). The P50 and the P95 are sufficient to fully characterize the lognormal distribution. Alternatively, it can be fully characterized by the combination of the P05 and the P95.

Generic uncertainty distributions may be used as default distributions in probabilistic hazard characterization, but also for deriving “data-based” single-value assessment factors. Thus, these factors could be used in non-probabilistic uncertainty analysis of a given hazard characterization. However, these empirical bounds (more exactly: the one on the conservative side) could differ from the usual default assessment factors.

In the following five subsections, the most frequently occurring aspects of hazard characterization are discussed consecutively. For each aspect, the aim is to translate relevant results from published studies into estimates of the GM and GSD of the uncertainty distribution of interest. The underlying data and the considerations in extracting information from these data are discussed in Annexes 3 and 4. Below, only the results are reported – that is, the uncertainty distributions that would appear to be most appropriate given the database used. As already noted, these may be considered as preliminary distributions. However, although they are not the result of a systematic review, it could be argued that they are based on data, which is not always the case for the currently used default factors.

## 4.2 Points of departure

The PoD is based on a specific data set, so the uncertainties in the PoD derive from the uncertainties in that data set. In this section, we discuss uncertainties with respect to two approaches to specifying the PoD from a chemical-specific dose–response data set: the BMD approach and the NOAEL approach. The NOAEL approach may result in a NOAEL or in a lowest-observed-adverse-effect level (LOAEL), and the uncertainties associated with either of these need to be considered separately. Further, both NOAEL and LOAEL need to be discussed separately for continuous and for quantal end-points.

### 4.2.1 *Benchmark dose*

In the BMD approach, the uncertainties in the data set can be evaluated directly, resulting in a confidence interval for the BMD (i.e. the combination of the lower [BMDL] and upper [BMDU] confidence limits of the BMD), or by deriving a full uncertainty distribution for the BMD – for example, by the bootstrap method or by a Bayesian approach. This uncertainty distribution specifically holds for the chemical considered, and there is no need to consider generic uncertainties here. The uncertainties have already been evaluated in a chemical-specific way. Therefore, deriving a generic uncertainty distribution for the BMD is not relevant.

However, the uncertainties reflected by a BMD confidence interval may not cover all uncertainties involved. For instance, when the confidence interval is based on a single selected model, model uncertainty is not covered. This can be addressed by combining the results from various models resulting in a good fit (see, for example, EFSA, 2009). Another example is when data are aggregated, such as combining developmental data from different litters, so that litter effects cannot be accounted for. Finally, when human epidemiological data are used, there is often substantial uncertainty in exposure estimates that may not be accounted for in the derivation of modelled BMDs.

### 4.2.2 *No-observed-adverse-effect level*

When the PoD is quantified in terms of a NOAEL, the uncertainty in that value cannot be quantified in a case-specific sense. All that can be done is to try to capture the uncertainty in

NOAELs in a generic way. As the discussion below shows, the uncertainty in a NOAEL may be substantial for typical dose–response data. Therefore, it is emphasized that in all situations where dose–response data are available (possibly in the form of summary statistics), the uncertainties in the dose–response data can be accurately and case-specifically quantified by applying the BMD approach. This includes data sets for which no NOAEL could be identified (“LOAEL only”), discussed in the next section. Therefore, generic uncertainty distributions for the NOAEL are useful only in specific situations – for example, where the underlying dose–response data were not available or missed essential information (e.g. only means but no group sizes or standard deviations [SDs] are given). Another reason may be that it would take substantial effort to obtain the dose–response data, and it may be desirable to first roughly estimate the margin of exposure, which might be large enough to easily cover the uncertainties associated with the NOAEL. However, as even the uncertainties around a NOAEL are uncertain, it is better to avoid the NOAEL whenever it is not really needed.

The NOAEL may be regarded as a rough estimate of the  $BMDL_x$ , where  $x$  is the default BMR (see EFSA, 2009). Thus, the generic uncertainty in the NOAEL may be defined as the precision of the NOAEL in estimating the BMDL. Useful historical data can be found in studies that compare NOAELs with BMDLs in the same data sets: the distribution of the NOAEL/BMDL ratios reflects the uncertainty in the NOAEL and can be used to inform a generic uncertainty distribution for the NOAEL. For the purpose of characterizing uncertainty in the framework presented in section 3, the uncertainty in the NOAEL must be translated into uncertainty in the associated (hypothetical) BMD rather than in the BMDL. Therefore, two steps are involved: (1) assessing the uncertainty in the NOAEL as an estimate of the BMDL (based on historical data); and (2) translating this uncertain BMDL into the uncertainty in the (true) target BMD (by assuming a 9-fold distance between the BMDL and BMDU). Combining these two steps results in the distributions proposed in Table 4.1 (see Annex 3 for more details on their derivation). Note that the P50 is smaller than 1 in all cases, owing to the second step of extrapolating the BMDL to the midpoint of the BMD confidence interval.

**Table 4.1: Uncertainty in the BMD when using a NOAEL ( $AF_{PoD-NOAEL}$ )<sup>a</sup>**

<i>Type of end-point</i>	<i>Type of study (route)</i>	<i>P50</i>	<i>P95/P50</i>	<i>(P05, P95)</i>	<i>Source</i>
Continuous	Chronic or subchronic (oral)	1/3	4.7	(0.07, 1.6)	Based on Bokkers & Slob (2007) analysis, with a BMR = 5%.
Continuous	Developmental (oral)	1/3	7.0	(0.05, 2.3)	Based on Allen et al. (1994) analysis, with a BMR = 5%.
Quantal (deterministic)	Developmental (oral)	2/9	5.0	(0.04, 1.1)	Based on Allen et al. (1994) analysis, adjusted to an ED <sub>50</sub> .
Quantal (stochastic)	Developmental (oral)	2/3	4.7	(0.14, 0.32)	Based on Allen et al. (1994) analysis, with a BMR = 10%.

$AF_{PoD-NOAEL}$ : assessment factor for use of a NOAEL as the point of departure; BMD: benchmark dose; BMR: benchmark response; ED<sub>50</sub>: median effective dose; NOAEL: no-observed-adverse-effect level; P05: 5th percentile; P50: 50th percentile; P95: 95th percentile

<sup>a</sup> For continuous data, the BMD relates to a per cent change of 5%, for quantal data, to an extra risk of 10%.

For other study types, no useful data were found. With this lack of information, use of the same P50 value as for one of the above distributions, but with a larger P95/P50 ratio, might be considered to reflect the additional uncertainty.

#### **4.2.3 Lowest-observed-adverse-effect level**

At first sight, the uncertainty in use of a LOAEL to estimate a BMDL could be assessed using the same approach as for the other aspects, by considering the distribution of the LOAEL to BMDL ratios based on historical data. However, in this case, the approach is not meaningful. The distance between a LOAEL and BMDL (or NOAEL) strongly depends on the (true) effect size at the LOAEL. For example, when the LOAEL relates to a close to 100% observed incidence, the BMDL could be any value below the LOAEL – that is, the uncertainty is in fact infinite (note that a dose of zero is an infinitely low dose). Therefore, an isolated LOAEL without considering the underlying dose–response information cannot be used as a PoD.

In those cases where the lowest dose in a study has been reported as a LOAEL, the following options exist:

- The dose–response data are or can be made available, in terms of summary data – that is, including group sizes and, in the case of continuous data, the SDs or standard errors of the mean (SEMs). If so, the BMD approach should be applied to the data. The resulting BMD confidence interval will probably be wide – in some cases acceptably wide, in others extremely wide, possibly with a BMDL that is “zero”. The latter situation will occur with a large effect size at the lowest administered dose. A BMDL close to zero means that the data set does not inform the BMD (at the specified BMR). One may then either increase the BMR and check whether that would result in an acceptable confidence interval for the associated BMD or reject the study as a basis for (quantitative) hazard characterization.
- The dose–response data are incomplete – for example, observed responses are available but lack the group sizes or the SDs/SEMs associated with the (continuous) group mean. In this case, one could semi-quantitatively guess about the possible shapes of the true dose–response relationship. Indeed, there is an element of guessing involved, as the uncertainty in the observed responses cannot be assessed due to the lack of information on group sizes or SDs/SEMs.
- There is no (or hardly any) dose–response information. In this case, the LOAEL cannot be used as a PoD.

It should be noted that extrapolating a LOAEL to a NOAEL without considering the underlying dose–response data is not warranted for the same reason: the effect size at the LOAEL could be very large, and the distance between the LOAEL and NOAEL (i.e. the LOAEL to NOAEL ratio) could have any value. Various studies have reviewed historical data with the purpose of establishing the distribution in the LOAEL/NOAEL ratio. This distribution, however, predominantly reflects the distribution of the dose spacing used in toxicological studies.

### **4.3 Exposure duration**

The uncertainties in many other aspects may be informed by distributions of observed ratios of PoDs, related to the relevant studies. For instance, a subchronic PoD may be regarded as a

surrogate or estimate of the chronic PoD, and the distribution of observed subchronic to chronic PoD ratios will inform the associated uncertainty at a generic level. It should be noted that PoD ratios in the review studies may be reported either as NOAEL or as BMD ratios, and both may be used for informing the uncertainty in a particular hazard characterization aspect. Before discussing a number of aspects consecutively, it is important to keep in mind the following two general principles, which apply equally to all extrapolation aspects discussed below:

1. BMD ratios rather than BMDL ratios are the relevant ones for the present purpose (as opposed to the previous discussion on the uncertainty distribution for the NOAEL, where NOAEL/BMDL ratios were needed). The reason is that we are interested in the distance between the “true” BMDs – for example, related to subchronic versus chronic exposure – and the observed BMDs are “best” estimates of the true BMDs. The BMDL is a conservative estimate, which has already taken the uncertainties in the (specific) dose–response data into account.
2. NOAEL ratios are found to show wider distributions than BMD ratios. This follows directly from the fact that the uncertainty associated with NOAELs is greater than for BMDs. If the uncertainty in the NOAEL has been separately accounted for by a generic distribution, as suggested in the previous section, then, also for the NOAEL, the uncertainty in a given aspect can be covered with an uncertainty distribution that is based on BMD ratios. Otherwise, the uncertainty in the NOAEL would be “double-counted”.

The results of the evaluation (described in Annex 3) are presented in Table 4.2.

**Table 4.2: Uncertainty in the chronic BMD when using a shorter-duration BMD ( $AF_{Dur}$ )**

<i>Type of duration extrapolation</i>	<i>P50</i>	<i>P95/P50</i>	<i>(P05, P95)</i>	<i>Source</i>
Subchronic to chronic	2	4	(1/2, 8)	Based on Bokkers & Slob (2005) analysis of BMD ratios in oral studies, but consistent with multiple analyses of NOAEL ratios in multiple species by both oral and inhalation exposures.
Subacute to chronic	5	8	(5/8, 40)	Estimated based on multiple analyses of data on NOAELs in multiple species by both oral and inhalation exposures.

$AF_{Dur}$ : assessment factor for exposure duration; BMD: benchmark dose; NOAEL: no-observed-adverse-effect level; P05: 5th percentile; P50: 50th percentile; P95: 95th percentile

#### 4.4 Interspecies extrapolation

In general, interspecies extrapolation may be subdivided into two parts:

1. adjustment of the dose for (generic) differences in body size between test animals and humans; and
2. accounting for potential remaining (chemical-specific) differences in toxicokinetics and toxicodynamics (denoted as TK/TD differences below).

Similar to duration extrapolation, PoD ratios are used to estimate the uncertainty distribution for interspecies adjustment.

#### 4.4.1 Body size adjustment

The first step is assumed to depend on the body sizes of the two species involved, but not on the specific chemical. Different approaches are used for oral and inhalation routes of exposure.

##### 4.4.1.1 Adjustment of oral doses

One way to adjust oral doses for body size differences between test animals and humans is to divide the applied dose (or PoD) by the body weight of the animal. Based on theoretical arguments, it has been argued that the dose should instead be divided by body weight to some power (“allometric scaling”) (e.g. USEPA, 2011). The value of that power is somewhat uncertain, however. Some have proposed a value of  $\frac{2}{3}$ , representing “surface area” scaling; others have proposed a power of  $\frac{3}{4}$ , representing metabolic adjustment. Various empirical studies are consistent with allometric scaling, but the data are not precise enough to distinguish between the two theoretical values: in general, they estimate the power to lie somewhere between 0.66 and 0.74 (e.g. Bokkers & Slob, 2007). Assuming a normal uncertainty distribution with a mean of 0.7 for the power, a SD of 0.024 would cover this uncertainty.

##### 4.4.1.2 Uncertainty distribution for allometric scaling factor for oral doses

To change a dose per kilogram body weight (bw) into an allometrically scaled dose, the following allometric factor needs to be applied:

$$AF_{\text{Inter-BS}} = \text{allometric factor} = \frac{\alpha \text{ bw}_{\text{human}}^{\alpha}}{\alpha \text{ bw}_{\text{test species}}^{\alpha}} \quad (4-1)$$

where  $AF_{\text{Inter-BS}}$  is the interspecies assessment factor for body size differences and  $\alpha$  is assumed to have a normal uncertainty distribution with a mean of 0.7 and a SD of 0.024. It follows that the allometric factor is lognormally distributed, with:

$$P50 = \frac{\alpha \text{ bw}_{\text{human}}^{\alpha}}{\alpha \text{ bw}_{\text{test species}}^{\alpha}} \quad ; \quad P95/P50 = \frac{\alpha \text{ bw}_{\text{human}}^{\alpha}}{\alpha \text{ bw}_{\text{test species}}^{\alpha}} \quad (4-2)$$

where  $\sigma$  denotes the SD of  $\alpha$  (0.024 in this example) and  $z$  is the 95th percentile of the standard normal distribution ( $z$ -score) (of 1.64).

As an example, suppose the test species was a rat with a body weight of 400 g and the target human being has a body weight of 60 kg; then the uncertainty distribution of the allometric factor has a  $P50 = 150^{0.3} = 4.5$  and a  $P95/P50 = 150^{1.64 * 0.024} = 1.2$ .

##### 4.4.1.3 Body size adjustment of inhalation exposures and its uncertainty

For inhalation exposures, different types of body size assessment factors have been derived for particles (regional deposited dose ratio, or RDDR) and gases (regional gas dose ratio, or RGDR). As defined in USEPA (1994), these factors are ratios of the equivalent exposure concentrations in test animals and humans, based on interspecies information about respiratory tract geometries and air flow rates, and differ depending on whether the effects of interest are regional or systemic. For example, effects in the upper airways are based on the

surface areas of relevant regions of the respiratory tract and the inhalation minute volume, whereas systemic effects that involve transport by blood utilize information on blood–air and blood–tissue partition coefficients. Thus, the P50 value for interspecies differences in inhalation exposures would be case specific and equal to  $AF_{\text{Inter-BS}} = 1/RDDR$  or  $1/RGDR$ . For gases, the RGDR is often assumed to be 1, which corresponds to the case of a systemically acting gas where the animal blood:air partition coefficient is greater than or equal to the human blood:air partition coefficient. Although there has not been a formal evaluation of uncertainty in the derivation of these factors, it may be assumed that the P95/P50 ratio is no more than 2, because these factors are largely based on physiological information.

#### 4.4.2 Toxicokinetic/toxicodynamic differences

Body size adjustment accounts for generic body size–related physiological differences that have been shown to be accurate on average across chemicals. Any remaining TK/TD differences between species depend on the specific chemical, but they will be described by an overall distribution. Thus, the TK/TD distribution describes the variation among chemicals in TK/TD differences between the two species (after adjustment for body size). When considering a specific chemical, the TK/TD difference for that chemical should be a particular value in the TK/TD distribution, but it is unknown which one. Therefore, the TK/TD distribution can be used as an uncertainty distribution for the TK/TD difference of any single chemical for which no chemical-specific or chemical class–specific TK or TD data are available. If chemical (class)–specific data are available, a chemical-specific uncertainty distribution for the TK/TD difference may be developed (rather than a single-value CSAF).

In principle, the TK/TD distribution could be separated into a TK distribution and a TD distribution, if sufficient data were available to inform them separately. Such data, however, have not been reviewed to the extent that separate uncertainty distributions could be derived. Also note that in a higher-tier hazard characterization that uses a PBTK model for interspecies extrapolation, allometric scaling is in fact taken into account in scaling the organ sizes and the metabolic rates in the model.

The results of the review and evaluation of available data (described in Annex 3) are presented in Table 4.3.

**Table 4.3: Uncertainty in interspecies BMD ratio due to remaining (chemical-specific) TK/TD differences after adjusting to account for body size differences ( $AF_{\text{Inter-TK/TD}}$ )**

<i>Type of study</i>	<i>P50</i>	<i>P95/P50</i>	<i>(P05, P95)</i>	<i>Source</i>
Subchronic or chronic	1	3	(1/3, 3)	Based on Bokkers & Slob (2007) analysis of BMD ratios for the same end-point in oral studies

$AF_{\text{Inter-TK/TD}}$ : interspecies assessment factor for remaining toxicokinetic and toxicodynamic differences; BMD: benchmark dose; P05: 5th percentile; P50: 50th percentile; P95: 95th percentile; TK/TD: toxicokinetic/toxicodynamic

## 4.5 Human interindividual differences in sensitivity

### 4.5.1 Background on uncertainties in human interindividual differences

Addressing the uncertainties related to the aspect of intraspecies differences in sensitivity differs from the other aspects in the sense that the uncertainty relates to variability in equipotent doses rather than a ratio of two specific equipotent doses (e.g. between median animal and median human being). Intraspecies variability may be informed by historical data that allow estimation of the variability in equipotent doses among individuals exposed to the same compound. Then, the variation among the individual equipotent doses for the same chemical can be calculated as the GSD, which will be denoted as  $GSD_H$ , where H stands for human.

By doing this for a set of chemicals, a set of  $GSD_H$  will be obtained. Based on a review of available data (Hattis & Lynch, 2007), the variation among chemicals regarding the interindividual variability in equipotent doses may be estimated. In particular, the set of  $\log(GSD_H)$  values may be described as a lognormal distribution with its own GSD that may serve as a measure of uncertainty, analogous to the other aspects. It will be denoted as  $GSD_U$ , where U stands for uncertainty.

In the assessment of any specific chemical (without specific human variability information), it is assumed that this chemical is a random draw from the same population of chemicals that comprised the database underlying the estimated variation in interindividual variability. The latter uncertainty distribution will, as for the other aspects, be characterized by the combination of P50 and P95 or by the combination of P05 and P95 of the distribution of interindividual variability estimates.

The distinction between  $GSD_U$  and  $GSD_H$  is further illustrated in Fig. 4.1. Care needs to be taken not to confuse the two:  $GSD_U$  is used for the uncertainty in  $GSD_H$ , the interindividual variation in equipotent doses. Further, care needs to be taken regarding the use of logs. As equipotent doses are assumed lognormal,  $GSD_H$  relates to the (lognormally distributed) equipotent dose, whereas  $GSD_U$  relates to the (lognormally distributed) logarithm of  $GSD_H$ .

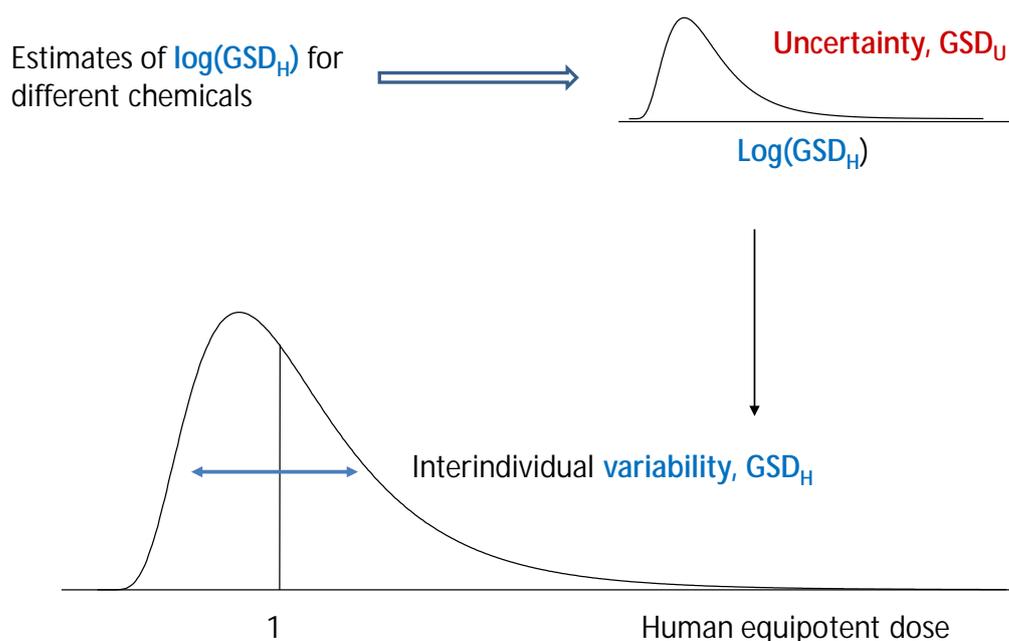
As discussed in section 3.1, for a particular hazard characterization, variability needs to be expressed in terms of the target incidence  $I$  in the population. Thus, an incidence  $I$  of, say, 5% corresponds to 95% of the population being protected against the specified effect  $M$ . For a given incidence, the factor needed to cover the corresponding fraction of the population is calculated using the  $GSD_H$  and the corresponding  $z$ -score of the normal distribution, as follows:

$$AF_{\text{Intra-}I} = \text{Factor covering } (1 - I) \text{ of the population} = GSD_H^{z_{1-I}} \quad (4-3)$$

For  $I = 5\%$ , 1% and 0.1%, the corresponding values for  $z_{1-I}$  are 1.6449, 2.3263 and 3.0902.

### 4.5.2 Uncertainty distributions for intraspecies variability

It is useful to further split up the equipotent dose distribution into two sub-distributions, reflecting the two portions of the causal pathway between external dose and end effect, as follows:



**Fig. 4.1: Distinction between the distribution reflecting the variability in human equipotent doses and the distribution reflecting the uncertainty in that variability, resulting from differences among chemicals in intraspecies variation.** Note that human equipotent doses are assumed to be lognormally distributed, whereas  $GSD_U$ , the uncertainty in  $\log(GSD_H)$ , is assumed to be lognormally distributed as well.

1. *Toxicokinetic variability* – characterized as  $GSD_{H-TK}$  and defined as differences among people in the external dose required to produce a similar systemic internal dose (concentration–time combination for systemically acting agents), usually measured in the blood<sup>14</sup>; and
2. *Toxicodynamic variability* – characterized as  $GSD_{H-TD}$  and defined as differences among people in the internal dose required to produce an effect of defined degree or severity ( $M$ ).

Breaking up the causal chain in this way can be helpful, because chemical-specific TK data are usually easier to obtain than TD data. In such cases, it may be possible to substitute chemical-specific information for the TK portion of the pathway, while retaining assumptions based on generic data for the TD portion, as has commonly been done in the IPCS procedure for deriving CSAFs (IPCS, 2005).

The available data are reviewed in Annex 4, from which uncertainty distributions for  $\log(GSD_{H-TK})$  and for  $\log(GSD_{H-TD})$  can be derived. Note that for deriving the uncertainty distribution for TD, the data have been restricted to systemic, non-immune-mediated effects. The results of the evaluation are presented in Table 4.4. Additionally, values for the intraspecies assessment factor  $AF_{Intra-I}$  calculated using equation 4-3 are presented in Table 4.5. Table 4.5 also presents the results of approximating the distribution of  $AF_{Intra-I}$  by a lognormal distribution, as is done by the APROBA tool. Clearly, the approximation grows worse with smaller values of incidence  $I$ .

<sup>14</sup> As discussed in Annex 4, many of the studies that are the basis of the TK distributions were originally phase I drug studies and therefore could be classified as “pharmacokinetics” rather than “toxicokinetics”. However, the term TK is used here to include these results to avoid confusion and also because it is thought that the interindividual variability in pharmacokinetic parameters observed in those studies is likely to be similar to that which would be observed with toxic chemicals that are the subjects of risk assessment evaluations.

**Table 4.4: Uncertainty distributions for intraspecies variability<sup>a</sup>**

<i>Parameter</i>	<i>P50</i>	<i>P95/P50</i>	<i>(P05, P95)</i>	<i>Source</i>
Log(GSD <sub>H-TK</sub> )	0.167	2.43	(0.0687, 0.407)	Based on AUC variability from oral exposures
Log(GSD <sub>H-TD</sub> )	0.221	2.85	(0.0776, 0.631)	Based on observations of systemic, non-immune-mediated, continuous physiological parameter changes or quantal biological response in relation to internal measures of systemic exposures
Log(GSD <sub>H</sub> )	0.324	2.152	(0.151, 0.697)	Based on Monte Carlo simulation combining log(GSD <sub>H-TK</sub> ) and log(GSD <sub>H-TD</sub> ), assuming independent lognormal distributions <sup>b</sup>

AUC: area under the concentration–time curve; GSD: geometric standard deviation; GSD<sub>H</sub>: interindividual variability in the human equipotent dose distribution; GSD<sub>H-TD</sub>: toxicodynamic variability in the human equipotent dose distribution; GSD<sub>H-TK</sub>: toxicokinetic variability in the human equipotent dose distribution; P05: 5th percentile; P50: 50th percentile; P95: 95th percentile

<sup>a</sup> To estimate the factor associated with a specific population incidence *I*, the distribution for log(GSD<sub>H</sub>) needs input into equation 4-3 (see Table 4.5).

<sup>b</sup> Specifically, a Monte Carlo simulation was performed, drawing independent samples from log(GSD<sub>H-TK</sub>) and log(GSD<sub>H-TD</sub>), such that the total variation  $\log(\text{GSD}_H)^2 = \log(\text{GSD}_{H-TK})^2 + \log(\text{GSD}_{H-TD})^2$ . Then the distribution for log(GSD<sub>H</sub>) was fitted by a lognormal distribution.

**Table 4.5: Uncertainty distributions for AF<sub>Intra-I</sub> for intraspecies variability for selected values of population incidence *I***

<i>Incidence (%)</i>	<i>Values based on Table 4.4 and equation 4-3</i>			<i>Lognormal approximation<sup>a</sup></i>	
	<i>P50</i>	<i>P95/P50</i>	<i>(P05, P95)</i>	<i>P50<sub>approx</sub></i>	<i>P95/P50<sub>approx</sub></i>
10	2.60	3.01	(1.56, 7.83)	3.49	2.24
5	3.41	4.11	(1.77, 14.02)	4.98	2.82
1	5.67	7.39	(2.24, 41.88)	9.69	4.32
0.50	6.83	9.15	(2.44, 62.52)	12.36	5.06
0.10	10.03	14.23	(2.92, 142.78)	20.42	6.99
0.05	11.64	16.92	(3.13, 196.93)	24.82	7.93
0.01	16.03	24.44	(3.63, 391.81)	37.71	10.39

P05: 5th percentile; P50: 50th percentile; P95: 95th percentile

<sup>a</sup> Lognormal approximation used in APROBA tool approximates the distribution based on Table 4.4 and equation 4-3 using a lognormal distribution that has the same values of P05 and P95 (i.e. by setting  $P50_{approx} = (P05 \times P95)^{1/2}$ ). Therefore, although the approximation has a different P50, the 90% confidence interval is the same.

The interpretation and use of these data must come with several caveats:

- On one hand, there has been no effort in the analysis of these data to remove the effects of measurement errors, including estimation errors (related to the equipotent doses), sampling errors (due to limited sample sizes) and errors in the assumed dose–response relationships. Measurement errors undoubtedly have spread the observations of individual parameter values farther apart than they are in reality, leading to a tendency to overestimate all the GSD<sub>H</sub> values summarized here. Similarly, this holds for the estimated variation among chemicals (P95/P50 ratios).
- On the other hand, the populations studied by the original investigators undoubtedly were less diverse than the general human population or subpopulations whose risks

are of interest for health protection from chemical exposures. Phase I drug studies, the source of much of the TK data summarized here, tend to include primarily healthy adults (often by design).

- The databases for both TK and TD variability have primarily been constructed from observations related to drugs. Drugs might tend to differ in both chemical properties and mode of action from substances that are considered for regulation as environmental or food contaminants. Drugs will often be more water soluble and be designed to act directly on specific macromolecular receptors without a need for metabolic activation. Environmental chemicals may more frequently tend to act on the body in less specific ways and require metabolism to generate forms with greater biological activity. The extent of these possible differences between ensembles of drugs and environmental chemicals and implications for TK and TD variability have not yet been quantitatively assessed.

#### 4.6 Route-to-route extrapolation

There is very limited empirical information comparing PoDs for different routes of exposure. The data that are available show a wide range of ratios of equipotent doses per unit body weight (see Annex 3).

In some cases, simplified biological modelling can be used to derive theoretical relationships between routes. One example is for systemic-acting volatile organic compounds, in which steady-state TK can be used (Chiu & White, 2006). In this case, the extrapolation from oral to inhalation uptake (or vice versa) can be expressed as follows:

$$\begin{aligned} & \text{Inhalation concentration} \times \text{Alveolar ventilation rate} \times \text{Inhalation fraction absorbed} = \\ & \text{Oral dose} \times \text{Oral fraction absorbed} \times \text{RtR factor} \end{aligned} \quad (4-4)$$

where RtR refers to the route-to-route extrapolation factor.

If the toxic moiety is a metabolite formed in the liver or is the parent compound when the target is the liver, then the RtR factor is given by:

$$\text{RtR factor}_{\text{Metabolism}} = 1 + \text{Lung clearance} / \text{Liver blood flow} \quad (4-5a)$$

$$\text{Lung clearance} = \text{Alveolar ventilation rate} / \text{Blood-air partition coefficient} \quad (4-5b)$$

On the other hand, if the toxic moiety is the parent compound for a target tissue other than the liver, then the RtR factor is given by:

$$\text{RtR factor}_{\text{Parent}} = 1 / (1 + \text{Liver clearance} / \text{Liver blood flow}) \quad (4-6a)$$

$$\text{Liver clearance} = \text{First-order clearance for hepatic metabolism} (= V_{\max}/K_m) \quad (4-6b)$$

where  $V_{\max}$  is the maximum initial rate of an enzyme-catalysed reaction and  $K_m$  is the substrate concentration at which the initial reaction rate of an enzyme-catalysed reaction is half maximal.

Notably, the above formula still requires reliable data on absorbed fractions, ideally matched for the concentration dependency (including saturation) and the temporal pattern of absorption.

Overall, the uncertainty in route-to-route extrapolation may be very large in the generic case in which little or no TK information is available and no reliable historical data have been located with which to construct preliminary uncertainty distributions. However, under specific conditions in which the effects are systemic and there is some chemical-specific information available, the uncertainty may be reduced substantially.

#### 4.7 Summary of generic uncertainties per aspect

Table 4.6 summarizes the conclusions with respect to generic uncertainties per aspect assuming lognormal uncertainty distributions (as used by the APROBA tool). Note that for uncertainties associated with using a LOAEL and with route-to-route extrapolation, no generic uncertainties could be ascertained. Additionally, the uncertainty in the BMD is case specific, derived from BMD analysis of the specific data set.

For intraspecies variability, the uncertainties depend on the target incidence level  $I$  and reflect an approximation for the uncertainty in  $AF_{\text{Intra-}I}$  as a lognormal distribution. Factors are provided for incidences of 5%, 1% and 0.1%, which correspond to the factor by which the 95th, 99th and 99.9th percentiles of the population are more sensitive than the median individual (factors for additional values of  $I$  were presented in Table 4.5).

#### 4.8 Evaluating primary uncertainty for other aspects

So far, this section has discussed how uncertainties could be quantified for the aspect PoD and for the following extrapolation aspects: interspecies, intraspecies, exposure duration and route-to-route. Some of the other aspects mentioned in Table 3.1 (section 3.1) might be evaluated in analogous ways. For instance, missing study uncertainty could be evaluated using PoD ratios related to two study types for a set of chemicals. When one of these two studies is missing in a specific case, the PoD ratio distribution may be used to reflect the associated uncertainty. For instance, the review studies by Janer et al. (2007a, 2007b) provide some information on PoD ratios, comparing the two-generation study against the one-generation studies and against the subchronic study, respectively.

Given the general observation that PoD ratios tend to closely follow a lognormal distribution, any uncertainty that is quantified in terms of two values (e.g. P50 and P95/P50 or P05 and P95) can directly be translated into an uncertainty distribution. In this way, these uncertainties can be included in the APROBA tool, where various cells are designated to be filled in by user-specified “other aspects” (note that in APROBA, the P05 and P95 are called LCL and UCL, for lower and upper confidence limits).

There may also be uncertainties that are not directly amenable to quantification in the sense of distributions derived from data, as the necessary data may be lacking. An example can be found in the deoxynivalenol case-study (see Annex 5), where the quality of the developmental study was considered an uncertain aspect. Based on expert judgement, the NOAEL in this study was considered to have been up to a factor of 5 higher than it should have been. Therefore, an uncertainty distribution was assumed with both P50 and P95/P50

equal to the square root of 5. By quantifying this uncertainty in this way, it could be included in the probabilistic analysis in the usual way.

**Table 4.6: Summary of generic uncertainties for different aspects of hazard characterization assuming lognormal uncertainty distributions**

Aspect of hazard characterization	Lognormal P50	Lognormal P95/P50	Lognormal (P05, P95)	Comments
<b>PoD uncertainty for NOAEL<sup>a</sup>: AF<sub>PoD-NOAEL</sub></b>				
Continuous end-point, chronic/subchronic study	1/3	4.7	(0.07, 1.6)	Ratio of NOAEL to BMD <sub>05</sub> (5% relative change)
Continuous end-point, developmental study	1/3	7.0	(0.05, 2.3)	Ratio of NOAEL to BMD <sub>05</sub> (5% relative change)
Deterministic quantal end-point	2/9	5	(0.04, 1.1)	Ratio of NOAEL to ED <sub>50</sub> (50% response)
Stochastic quantal end-point	2/3	4.7	(0.14, 3.2)	Ratio of NOAEL to BMD <sub>10</sub> (10% extra risk)
<b>Exposure duration: AF<sub>Dur</sub></b>				
Subchronic à Chronic	2	4	(1/2, 8)	–
Subacute à Chronic	5	8	(5/8, 40)	–
<b>Interspecies body size adjustment: AF<sub>Inter-BS</sub></b>				
Oral	$\frac{\frac{\sigma}{\mu} \frac{bw_{human}}{bw_{test\ species}}}{0.3}$	$\frac{\frac{\sigma}{\mu} \frac{bw_{human}}{bw_{test\ species}}}{0.04}$	$\frac{\frac{\sigma}{\mu} \frac{bw_{human}}{bw_{test\ species}}}{(0.26, 0.34)}$	Use case-specific body weights
Inhalation	1/RDDR or 1/RGDR	2	(0.5, 2)/RDDR or (0.5, 2)/RGDR	Use case-specific RDDR (particle) or RGDR <sup>b</sup> (gas)
<b>Interspecies TK/TD differences: AF<sub>Inter-TK/TD</sub></b>				
Oral	1	3	(1/3, 3)	Given lack of alternative, can also be used for inhalation
<b>Intraspecies differences for incidence I: AF<sub>Intra-I</sub></b>				
I = 5%	5.0	2.8	(1.8, 14)	Log(GSD <sub>H</sub> ) P50 = 0.32 and P95/P50 = 2.2
I = 1%	9.7	4.3	(2.2, 42)	
I = 0.1%	20.4	7.0	(2.9, 143)	

BMD<sub>x</sub>: benchmark dose for x% benchmark response; bw: body weight; ED<sub>50</sub>: median effective dose; GSD<sub>H</sub>: geometric standard deviation for interindividual variability in the human equipotent dose distribution; NOAEL: no-observed-adverse-effect level; P05: 5th percentile; P50: 50th percentile; P95: 95th percentile; PoD: point of departure; RDDR: regional deposited dose ratio; RGDR: regional gas dose ratio; TK/TD: toxicokinetic/toxicodynamic

<sup>a</sup> When using a NOAEL as the PoD, the uncertainty includes both the fact that the NOAEL is an approximation for the BMDL as well as the uncertainty in the underlying BMD (a ratio of 3 is assumed between the median estimate of the BMD and the BMDL).

<sup>b</sup> For gases, the RGDR is often assumed to be 1.

Distributions generated in this way – that is, by expert judgement rather than using historical data relevant to the substance and end-point at hand – might be perceived as being associated with a higher degree of secondary uncertainty; in other words, one can be less certain that the result of the primary uncertainty analysis is accurate (enough). Dealing with secondary uncertainties is treated below, in section 4.10.

#### 4.9 Chemical-specific distributions for primary uncertainties

In cases where a CSAF (IPCS, 2005) may be derived, it would be appropriate to try to quantify the uncertainty around that value by an uncertainty distribution. For instance, if, in a particular instance, animals are judged to be between 1- and 4-fold more sensitive than humans after allometric scaling, then the CSAF for  $AF_{inter-TK/TD}$  may be reflected by a distribution with  $GM = 1/2$  and  $P95/P50 = 2$  (as opposed to the suggested generic distribution based on historical data having  $GM = 1$  and  $P95/P50 = 3$ ; this also reflects that uncertainty has decreased due to chemical-specific knowledge).

Ideally, the derivation of the CSAF itself would include an uncertainty analysis of the model or the data on which the uncertainty distribution for the CSAF was based. For instance, if an oral interspecies CSAF were developed from a PBTK model, the deterministic approach would be simply to estimate the animal internal dose at the PoD and then calculate the human equivalent oral dose (HED) that corresponds to that same internal dose. Then the CSAF would equal the PoD/HED. However, to account for uncertainties, one must acknowledge uncertainty in both experimental animal and human internal dose estimates. Table 4.7 gives an example in which probabilistic PBTK models for trichloroacetic acid (TCA) were used to derive mouse and human internal dose estimates as a function of oral dose, along with their uncertainties (mouse TCA model: Chiu, 2011; human TCA model: submodel from trichloroethylene model of Chiu et al., 2009).

**Table 4.7: Example of deriving a chemical-specific uncertainty distribution for interspecies differences in toxicokinetics**

Row no.	Quantity	Units	P50	P95/P50
1	Mouse oral PoD (NOAEL)	mg/kg bw per day (mouse)	8	–
2	Mouse internal dose AUC at NOAEL	(mg·h)/(L·d)	56	2.50
3	Human internal dose / oral dose ratio <sup>a</sup>	(mg·h)/(L·d) per mg/kg bw per day (human)	81	1.35
4	Human oral dose at mouse PoD = Row 2 / Row 3	mg/kg bw per day (human)	0.69	2.62 <sup>b</sup>
5	$CSAF_{inter-TK} = \text{Row 1} / \text{Row 4}$	mg/kg bw per day (mouse) per mg/kg bw per day (human)	12	2.62

$CSAF_{inter-TK}$ : chemical-specific adjustment factor for interspecies toxicokinetics; NOAEL: no-observed-adverse-effect level; P50: 50th percentile; P95: 95th percentile; PoD: point of departure

<sup>a</sup> Assuming linear relationship in the internal dose range of interest. Estimate is for the typical (median) human.

<sup>b</sup> Combined assuming independent lognormal distributions.

Source: Chiu et al. (2009); Chiu (2011)

Note that the median estimate of the CSAF does not necessarily have to be smaller than the median from the general distribution – in fact, if the generic distribution is accurate, one would expect that, on average, half of the time the median CSAF would be greater than the generic median, and half of the time it would be less than the generic median. In this case, the generic interspecies scaling distribution (given an assumed mouse body weight of 0.035 kg and human body weight of 70 kg) has a P50 value of 9.8, so the P50 of the CSAF is slightly larger.

Additionally, in principle, the uncertainty in the CSAF distribution might be larger than the uncertainty in the generic distribution due to poor chemical-specific data. In that case, unless

there is little overlap between the CSAF and generic distributions, the generic distribution would likely be preferred, as it would appear that data from many chemicals are more informative than the available chemical-specific data.

Alternatively, in a fully Bayesian framework, the generic distribution could be considered as “a priori” information, and the CSAF distribution combined with it using Bayes’ theorem. The generic uncertainty distribution for body weight scaling has a P95/P50 ratio of 1.4, whereas the generic interspecies TK/TD distribution has a P95/P50 ratio of 3.0, with a combined P95/P50 ratio of 3.1. However, a direct comparison is complicated by the fact that the CSAF incorporates aspects that are related to TK only, and not TD.

This difficulty brings to light the additional issue with using CSAFs, of whether and how to modify (or remove) the generic uncertainty distributions of the various hazard characterization aspects in order to account for chemical-specific data. For intraspecies variability, the generic distributions described previously already separate TK and TD, so if a CSAF for intraspecies TK is developed, a TD-only generic distribution is available (and vice versa). However, such a division is not available for interspecies extrapolation. For instance, if an interspecies CSAF addresses only TK, then the body size scaling is removed (as a PBTK model already accounts for body size differences), but the generic interspecies TK/TD distribution is no longer fully applicable either (as a PBTK model at the same time accounts for TK). Removing  $AF_{\text{Inter-TK/TD}}$  entirely would imply that there is no remaining TD uncertainty, which is not likely to be true; however, retaining the original distribution would “double-count” the interspecies TK. A practical choice may be to assume equal, independent contributions from interspecies TK and TD and re-derive the generic distribution accordingly.

#### 4.10 Evaluating secondary uncertainties

The distributions derived in sections 4.1–4.9 may themselves be subject to uncertainty due to, for example, limitations in the quantity, quality or relevance of the data on which they are based. Table 4.8 summarizes elements potentially influencing secondary uncertainty for various aspects addressed in the generic hazard characterizations.

As explained in section 3.4.1, these secondary uncertainties make the overall (primary) uncertainty of the hazard characterization imprecise and need to be taken into account. There are two different conclusions that could be drawn from such an assessment in any individual case:

1. For none of the individual aspects of hazard characterization are secondary uncertainties likely to significantly impact the uncertainty analysis result. In other words, the secondary uncertainties are only “small” and are not expected to have a major impact on the primary uncertainty as evaluated in the probabilistic analysis. Yet the estimated (primary) uncertainty should be considered as a value with some degree of imprecision; or
2. For one or more aspects of hazard characterization, the extent to which secondary uncertainties are present is not small, and the overall uncertainty resulting from a probabilistic analysis should be considered as a rough estimate only and may merit further evaluation, as outlined below.

The rightmost column of Table 4.8 below aims to provide a judgement on secondary uncertainties in a generic way. However, for some of the aspects, a case-specific evaluation

might be needed. For example, secondary uncertainty might be considered to be low if the end-point under question is among those represented very well in the data set on which the respective generic uncertainty distribution was built. Conversely, secondary uncertainty might be significant if an effect is considered that was not at all covered in the database.

**Table 4.8: Description of secondary uncertainties related to a number of typical aspects in a hazard characterization**

<i>Aspect</i>	<i>Elements of secondary uncertainty in assumed primary uncertainty for the given aspect</i>	<i>Extent to which secondary uncertainty might impact primary uncertainty for the given aspect</i>
<b>A. Evaluating the dose–response data</b>		
BMDL available	Assuming that the BMD approach has been performed properly (taking model uncertainty into account), while the data contain only random sampling errors, the BMDL and BMDU should reflect the uncertainty adequately.	Negligible
NOAEL for continuous end-point	NOAEL uncertainty distribution reasonably reflects the uncertainty for liver weights, kidney weights and red blood cell counts in NTP studies. For other study types (and possibly other end-points), the underlying information is weaker.  For critical data sets of lesser quality or smaller group sizes, the distribution for NOAEL/BMDL ratios will be shifted to the right and will also be wider, but it is hard to say by how much.  The BMDU is assumed to be a factor of 9 higher than the BMDL, but this factor may be smaller in some data sets and larger in others.	Case specific
NOAEL for quantal end-points	The assumed NOAEL uncertainty distribution is mainly based on developmental end-points. For other types of effects or studies, the uncertainty distribution could differ in both GM and GSD, in both directions.  The BMDU is assumed to be a factor of 9 higher than the BMDL, but this factor may be smaller in some data sets and larger in others.	Case specific
Extrapolation from LOAEL, no dose–response data available	As the effect size at the LOAEL could be anything, primary uncertainty is unbounded; secondary uncertainty is infinite.	Unbounded
Extrapolation from LOAEL, dose–response data available, but no SDs or group sizes	When there are multiple doses showing a clear dose–response relationship, secondary uncertainty might be limited. In continuous data, using historical data for the within-group variation might help, if the group sizes are available.	Case specific
Extrapolation from LOAEL, dose–response data available, BMD approach applied after all	No major secondary uncertainties (see first row).	Negligible

**Table 4.8 (continued)**

<b>Aspect</b>	<b>Elements of secondary uncertainty in assumed primary uncertainty for the given aspect</b>	<b>Extent to which secondary uncertainty might impact primary uncertainty for the given aspect</b>
<b>B. Extrapolating between species</b>		
For oral dose: Uncertainty in allometric power is normal distribution with P05 = 0.6 and P95 = 0.8 (for oral dose)	Empirical evidence for allometric power to lie between the assumed values is strong.	Small
Adjusting for TK/TD differences (after allometric scaling): lognormal distribution with GM = 1 and P95/P50 = 3	The distribution is based on a limited number of end-points from repeated-dose studies and a possibly biased sample of chemicals (Bokkers & Slob, 2007). For other end-points, the distribution is not expected to have another GM, but it might be wider. However, if the interspecies differences are mainly driven by TK differences, the distribution for other end-points would be similar.	Case specific, but probably small in most cases
<b>C. Estimating intraspecies variability</b>		
Adjusting for interindividual differences in sensitivity	Estimation errors in $\log(\text{GSD}_H)$ may have resulted in fairly strong overestimation of both $\text{GM}_U$ and $\text{GSD}_U$ of the uncertainty distribution; in contrast, restricted study populations result in underestimation of both $\text{GM}_U$ and $\text{GSD}_U$ . Some simulations indicated a relatively small impact on the overall uncertainty.	Small
<b>D. Extrapolating subchronic to chronic exposure</b>		
Adjusting for exposure duration	Underlying data are relatively good; no major deviations are expected.	Small
<b>E. Other general sources of uncertainty</b>		
Use of lognormal distributions for all uncertainty factors	Distributions of observed ratios of PoDs are close to lognormal. Only the uncertainty distribution of a specific BMD is not always lognormal. The effect is hard to predict, but probably not dramatic.	Small
Probabilistic combination of uncertainty distributions	It is assumed that the distributions are independent of each other, which is a reasonable assumption.	Small
Use of approximation in APROBA tool	Avoids need for Monte Carlo simulations. Testing indicates that differences are small (see Annex 3).	Small

APROBA: Approximating PROBABILISTIC Analysis; BMD: benchmark dose; BMDL: lower confidence limit of the benchmark dose; BMDU: upper confidence limit of the benchmark dose; GSD: geometric standard deviation; GM: geometric mean; LOAEL: lowest-observed-adverse-effect level; NOAEL: no-observed-adverse-effect level; NTP: (United States) National Toxicology Program; P05: 5th percentile; P50: 50th percentile; P95: 95th percentile; PoD: point of departure; SD: standard deviation; TD: toxicodynamics; TK: toxicokinetics

It is difficult to estimate the quantitative impact of secondary uncertainty intuitively, without doing probability calculations. Therefore, a more reliable way of exploring the impact of secondary uncertainty on the outcomes from a probabilistic analysis is by a simple sensitivity analysis using the APROBA tool. With this tool, the confidence limits for the uncertainty distribution for a given aspect can be changed in an instant, and the impact on overall uncertainty of the  $\text{HD}_M^I$  can be studied very easily. For a practical example of such an

evaluation, the reader is referred to the case-study on deoxynivalenol, included as Annex 5 of this guidance document.

There are, of course, various results that could arise from such a sensitivity analysis. Alternative possible outcomes include the following:

- The impact of the secondary uncertainty on overall uncertainty could appear to be low and would not be regarded to affect the overall conclusion of the hazard (or risk) characterization.
- The impact of the secondary uncertainty on overall uncertainty could appear to be significant, but, given estimated exposure levels, this would be covered by a sufficient margin between the probabilistic RfD and the human exposure level.
- The impact of the secondary uncertainty on overall uncertainty could appear to be significant and also large enough to affect the overall conclusion of the hazard (or risk) characterization – that is, the margin between the probabilistic RfD and the estimated exposure level does not seem to cover that potential impact.

However, the evaluation of secondary uncertainties might also end with the conclusion that it was impossible to give a reasonable estimate of the impact of that uncertainty. In such a situation, if the result of the hazard characterization is communicated to risk managers, they should be alerted that the evaluated uncertainties are themselves subject to an unknown degree of uncertainty. Then, the only two logical alternatives for proceeding would be either to generate more reliable data that could reduce the secondary uncertainty for the particular aspect(s) or, if that is not an option, to make an appropriate regulatory decision in the light of uncertainty (and the paradigms and principles of the respective regulatory programme).

In any case, it is important to document the evidence and reasoning on which the assessment of these uncertainties is based. A tabular format may be helpful for summarizing this, as illustrated in the deoxynivalenol case-study (see Annex 5).

## 5. ILLUSTRATION OF THE APPROACH USING GENERIC HAZARD CHARACTERIZATION SCENARIOS

This section illustrates uncertainty characterization by applying the approach described previously to a number of generic hazard characterization scenarios. The results from an uncertainty analysis of a given hazard characterization do not depend on the value of the PoD itself, but rather on uncertainty in the PoD and on the assumed uncertainties related to the extrapolation steps involved. Therefore, generic results may be generated for typical hazard characterizations that use particular (default) assessment factors. As an illustration, a number of typical hazard characterizations were probabilistically analysed, and the results are presented below. These results are valid for similar hazard characterizations – that is, based on a PoD of the same type (NOAEL or BMDL) and using the same assessment factors, while the uncertainties are quantified by the same distributions. The examples below used the uncertainty distributions discussed in section 4.

Any analysis that is based on data will result in an outcome with some precision that may be small or large, depending on the quality of the data. Similarly, the outcome of a probabilistic analysis results in estimates of uncertainty, which are subject to some imprecision, due to the fact that the estimated uncertainties are subject to uncertainty as well. This is indicated by the term secondary uncertainty (see previous section for a more extensive discussion). It is important to realize that the uncertainty distributions underlying the uncertainty analysis are based on data, but these data are of varying quality. The overall uncertainties, as presented in the tables, should therefore be interpreted as approximate values only. Furthermore, it is reiterated that the numbers for protection goals and uncertainty measures as used in this document, including in the examples below, are meant for illustrative purposes only, and not as prescriptive numbers.

### 5.1 Probabilistic uncertainty characterization of the deterministic reference dose and derivation of a probabilistic reference dose

The following different scenarios are considered for both a continuous end-point and a quantal end-point: chronic or subchronic study, PoD = BMDL or NOAEL, and test animal is rat or mouse, comprising 16 scenarios in total. Results were generated with both the approximate probabilistic approach, implemented in the Microsoft Excel spreadsheet tool APROBA and a full probabilistic approach, using Monte Carlo simulation. Fig. 5.1 shows a screenshot of one of the scenarios.

Table 5.1 shows the results for the 16 scenarios for the estimated per cent coverage for “traditional” RfDs calculated by use of fixed 10-fold assessment factors, the estimated range of uncertainty, as well as probabilistic RfDs for specified coverage of 95%, assuming a 1% target population incidence ( $I = \text{sensitive} = 1\%$ ).

The main conclusions from the results in Table 5.1 are as follows:

- The per cent coverage in these hazard characterizations is fairly high, the lowest being 76%, as calculated by APROBA (83% by Monte Carlo), and most are over 90%. This confirms that a lower-tier hazard characterization might be considered to be reasonably conservative.

Evaluating and Expressing Uncertainty in Hazard Characterization

TITLE:		Sub-chronic Quantal (deterministic) Rat NOAEL	
<b>INPUTS RELATED TO STUDY, END-POINT AND PROTECTION GOALS</b>			
DESCRIPTION	INPUTS		COMMON VALUE(S)
End-point	Generic end-point		Case-specific
Data type	Quantal-deterministic		Case-specific
Data route	Oral		Case-specific
Study type	Subchronic		Case-specific
Test species	Rat		Case-specific
Body weight test species (kg)	0.400		0.4
Human median body weight (kg)	60.0		60
Target BMR (= <i>M</i> , user input for BMDLs only)	50%		50%
Population incidence goal (= <i>I</i> )	1%		5%, 1%, 0.1%, 0.01%
Probabilistic coverage goal	95%		95%
PoD type	NOAEL		Case-specific
PoD value	100		Case-specific
BMDU (User Input for BMDL PoDs)			Blank if PoD is NOAEL/LOAEL
PoD units	mg/kg body weight per day		mg/kg body weight per day
Deterministic overall AF	1000		Case-specific
Deterministic RfD	0.10		Calculated
Exposure estimate (optional)	1.0		User supplied
<b>INPUTS RELATED TO ADJUSTMENT, VARIABILITY AND UNCERTAINTY</b>			
HAZARD CHARACTERIZATION	INPUTS		PROVISIONAL VALUE(S)
PoD	LCL	100	Calculated from inputs
(Modelled BMD uncertainty)	UCL	100	Calculated from inputs
NOAEL to BMD	LCL	0.0444	0.04
(NOAEL or LOAEL only)	UCL	1.11	1.11
Interspecies scaling	LCL	3.68	3.68
(Allometric for oral)	UCL	5.49	5.49
Interspecies TK/TD	LCL	0.333	0.33
(Remaining TK & TD)	UCL	3.00	3.00
Duration extrapolation	LCL	0.500	0.50
	UCL	8.00	8.00
Intraspecies	LCL	2.24	2.24
	UCL	41.9	41.88
Other aspect #1	LCL	1.00	1.00
(Description here)	UCL	1.00	1.00
Other aspect #2	LCL	1.00	1.00
(Description here)	UCL	1.00	1.00
Other aspect #3	LCL	1.00	1.00
(Description here)	UCL	1.00	1.00
<b>NON-PROBABILISTIC ANALYSIS OUTPUTS <sup>1a</sup></b>			
Target Human Dose (HD <sub>M</sub> <sup>1</sup> )	LCL	0.0163	mg/kg body weight per day
	UCL	1638	mg/kg body weight per day
Fold Range of Uncertainty	100504		
Estimated "Coverage" of Non-Prob. LCL of HD <sub>M</sub> <sup>1*</sup>			100%
*Based on approximate probabilistic analysis, below.			
<b>APPROXIMATE PROBABILISTIC ANALYSIS OUTPUTS</b>			
Standard Confidence Interval			
Target Human Dose (HD <sub>M</sub> <sup>1</sup> )	LCL (P05)	0.311	mg/kg body weight per day
	UCL (P95)	85.9	mg/kg body weight per day
Degree of Uncertainty (Fold Range)			277
Estimated "Coverage" of Deterministic RfD			99%
Probabilistic RfD	= Approximate probabilistic HD <sub>M</sub> <sup>1</sup> at specified % confidence		
0.31	= Estimate of dose (mg/kg body weight per day) at which, with		
	95%	confidence	
	1%	of the population will have	Generic end-point

Fig. 5.1: APROBA screenshot for approximate probabilistic uncertainty characterization of a generic typical hazard characterization. The scenario for a subchronic NOAEL in a rat, for a deterministic quantal end-point, is illustrated here.

Table 5.1: Probabilistic uncertainty analyses for typical generic hazard characterizations<sup>a</sup>

Study duration/ type of end- point	Species	Type of PoD	Determin- istic RfD	Coverage of deterministic RfD <sup>b</sup>	Degree of uncertainty (P95/P05 of HD <sub>M</sub> <sup>01</sup> ) <sup>b</sup>	Probabilistic RfD (P05 of HD <sub>M</sub> <sup>01</sup> ) <sup>b</sup>
<b>Chronic</b>						
Continuous (e.g. per cent change in haematocrit):	Rat	BMDL <sup>c</sup>	1	<b>93%</b> [95%]	<b>73</b> [76]	<b>0.81</b> [0.94]
		NOAEL	1	<b>91%</b> [93%]	<b>123</b> [127]	<b>0.62</b> [0.75]
HD <sub>05</sub> <sup>01</sup>	Mouse	BMDL <sup>c</sup>	1	<b>78%</b> [86%]	<b>75</b> [78]	<b>0.32</b> [0.38]
		NOAEL	1	<b>76%</b> [83%]	<b>126</b> [130]	<b>0.25</b> [0.30]
Quantal- deterministic (e.g. mild histopathological lesion):	Rat	BMDL <sup>c</sup>	1	<b>93%</b> [95%]	<b>73</b> [76]	<b>0.81</b> [0.94]
		NOAEL	1	<b>94%</b> [96%]	<b>133</b> [137]	<b>0.90</b> [1.08]
HD <sub>mild</sub> <sup>01</sup>	Mouse	BMDL <sup>c</sup>	1	<b>78%</b> [86%]	<b>75</b> [78]	<b>0.32</b> [0.38]
		NOAEL	1	<b>83%</b> [88%]	<b>137</b> [141]	<b>0.36</b> [0.44]
<b>Subchronic</b>						
Continuous (e.g. per cent change in haematocrit):	Rat	BMDL <sup>c</sup>	0.1	<b>99%</b> [99%]	<b>165</b> [170]	<b>0.27</b> [0.33]
		NOAEL	0.1	<b>98%</b> [98%]	<b>257</b> [259]	<b>0.21</b> [0.27]
HD <sub>05</sub> <sup>01</sup>	Mouse	BMDL <sup>c</sup>	0.1	<b>95%</b> [96%]	<b>169</b> [174]	<b>0.11</b> [0.13]
		NOAEL	0.1	<b>94%</b> [96%]	<b>265</b> [266]	<b>0.086</b> [0.11]
Quantal- deterministic (e.g. mild histopathological lesion):	Rat	BMDL <sup>c</sup>	0.1	<b>99%</b> [99%]	<b>165</b> [170]	<b>0.27</b> [0.33]
		NOAEL	0.1	<b>99%</b> [99%]	<b>277</b> [279]	<b>0.31</b> [0.39]
HD <sub>mild</sub> <sup>01</sup>	Mouse	BMDL <sup>c</sup>	0.1	<b>95%</b> [96%]	<b>169</b> [174]	<b>0.11</b> [0.13]
		NOAEL	0.1	<b>96%</b> [97%]	<b>283</b> [286]	<b>0.13</b> [0.16]

BMDL: lower confidence limit of the benchmark dose; HD<sub>M</sub><sup>l</sup>: target human dose for magnitude of effect *M* and population incidence *l*; NOAEL: no-observed-adverse-effect level; P05: 5th percentile; P95: 95th percentile; PoD: point of departure; RfD: reference dose

<sup>a</sup> Based on a PoD (BMDL or NOAEL) of 100, with deterministic interspecies and intraspecies assessment factors of 10. For the subchronic duration, an additional deterministic assessment factor of 10 was included.

<sup>b</sup> Probabilistic analyses used assumed uncertainty distributions discussed in section 4 and a target population incidence *l* = 1%. **Bold** = results from APROBA; [square brackets] = results from Monte Carlo simulation with 20 000 samples.

<sup>c</sup> The 90% confidence interval of the BMD is assumed to be (100, 900). Both APROBA and Monte Carlo simulations assume that BMD is lognormally distributed.

- The degree of uncertainty – that is, the overall uncertainty in estimating the target human dose – can be very high, in particular when the PoD is a NOAEL (overall uncertainty range, i.e. P95/P05, up to a factor of almost 300 in some of these scenarios).
- Hazard characterizations with only interspecies and intraspecies assessment factors are less conservative than those with an additional subchronic assessment factor. This is due to the general phenomenon that the conservatism associated with multiplying factors increases with an increasing number of factors. It should be noted that a factor of 10 is used here for the additional assessment factor for subchronic to chronic extrapolation, which is higher than the default value used in some institutions. Obviously, the difference would have been smaller had a smaller subchronic assessment factor been used.
- Hazard characterizations based on a NOAEL result in a larger degree of uncertainty in all scenarios considered (due to the additional uncertainty in the NOAEL).
- Although the traditional RfD would not differ among the eight chronic scenarios or the eight subchronic scenarios, the probabilistic RfD does, reflecting the fact that the uncertainties in the various scenarios are not the same.
- Comparing the APROBA and Monte Carlo calculations, APROBA tends to underestimate coverage slightly. Accordingly, the probabilistic RfD from APROBA is slightly conservative, by being slightly (1.2- to 1.3-fold) lower (more strict) than that calculated using Monte Carlo simulation.

## 5.2 Sensitivity analysis: ranking the sources of uncertainty

An APROBA analysis also shows the contributions to overall uncertainty that can be attributed to each hazard characterization aspect. Fig. 5.2 shows a screenshot of the calculations showing the per cent contribution to uncertainty for each of the hazard characterization aspects applicable in this particular scenario.

For this scenario, the ranking from largest to smallest sources of uncertainty is:

1. use of a NOAEL as the PoD;
2. intraspecies variability (at  $I = 1\%$ );
3. duration extrapolation;
4. interspecies TK/TD differences; and
5. interspecies body size scaling.

This information can be used in the overall context of the question, “Do we know enough?”, discussed in section 2. Specifically, it identifies the areas of uncertainty that, if reduced, would have the greatest impact on the overall uncertainty, thereby helping to prioritize additional analysis or data generation. Put another way, it provides a “value of additional information” for each hazard characterization aspect in terms of its potential impact on overall uncertainty.

### 5.3 Visualization of the uncertainty in the target human dose for different values of coverage and population incidence

The analyses described above focus on characterizing the uncertainty (coverage) in the existing default approaches and deriving probabilistic RfDs for preset values of (1) magnitude  $M$ , (2) incidence  $I$  and (3) per cent coverage. However, the  $HD_M^I$  concept allows for consideration of multiple options with respect to all three of these components, enabling a richer quantitative characterization of hazard.

INTERMEDIATE CALCULATIONS FOR UNCERTAINTY ANALYSES				% contribution
ASPECT			$[\log(P95/P50)]^2$	to overall uncertainty
PoD	P50	100		--
	P95/P50	1.00	0	
NOAEL to BMD	P50	0.222		33%
	P95/P50	5.00	0.489	
Interspecies scaling	P50	4.50		1%
	P95/P50	1.22	0.00758	
Interspecies TK/TD	P50	1.00		15%
	P95/P50	3.00	0.228	
Duration extrapolation	P50	2.00		24%
	P95/P50	4.00	0.362	
Intraspecies	P50	9.69		27%
	P95/P50	4.32	0.404	
Other aspect #1 (Description here)	P50	1.00		--
	P95/P50	1.00	0	
Other aspect #2 (Description here)	P50	1.00		--
	P95/P50	1.00	0	
Other aspect #3 (Description here)	P50	1.00		--
	P95/P50	1.00	0	
Target Human Dose ( $HD_M^I$ )		<b>Non-Prob.</b>	<b>Approx. Prob.</b>	<b>Greatest contributor</b>
	P50	5.17	5.17	to overall uncertainty
	UCL/P50	317	16.6	NOAEL to BMD

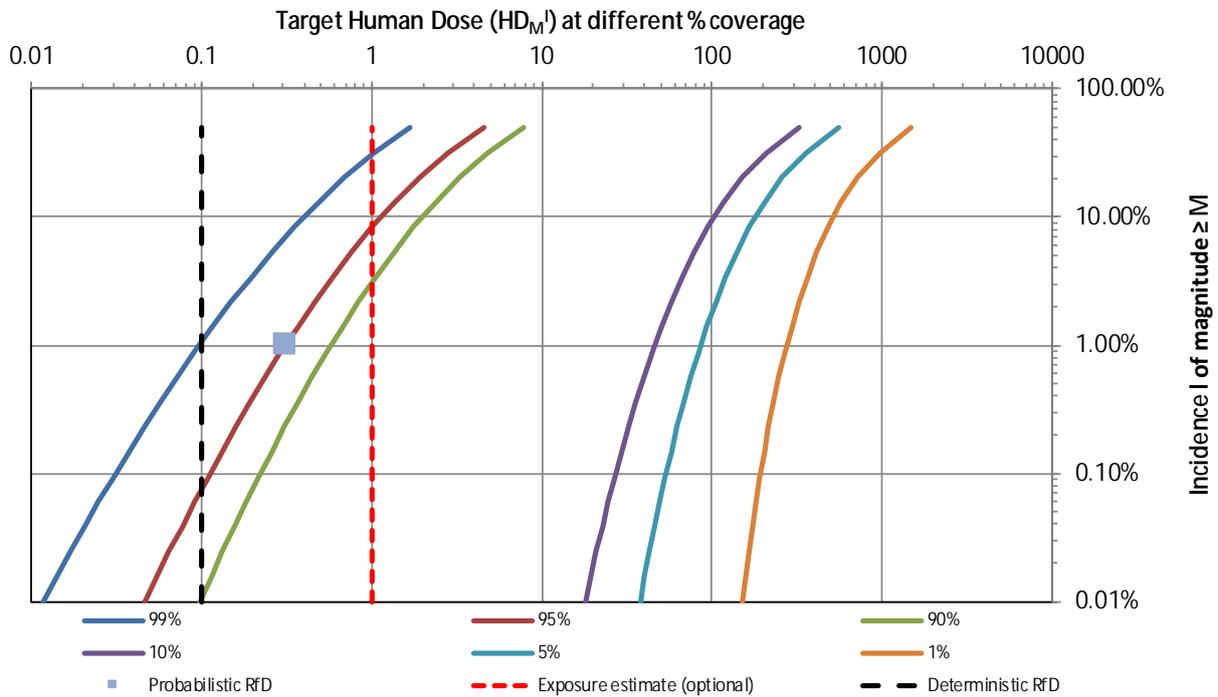
**Fig. 5.2: APROBA screenshot for characterizing the per cent contribution to overall uncertainty from each hazard characterization aspect.** The scenario for a subchronic NOAEL in a rat, for a deterministic quantal end-point, is illustrated here. Note: The approximate probabilistic analysis assumes that the intraspecies variability factor is lognormally distributed, with the same P05 and P95 as for the original distribution. Thus, the P50 and P95/P50 are not from the original distribution, but rather are calculated from the lognormal approximation.

The APROBA tool includes the ability to illustrate graphically the impact of different selected population incidences  $I$  and per cent coverage for a fixed magnitude of effect  $M$ . Fig. 5.3 shows an APROBA screenshot illustrating the relationship between the  $HD_M^I$  and population incidence  $I$  at different levels of coverage, providing a visual depiction of the impact of different choices of coverage and incidence on the target human dose.

Some features evident from this particular case include the following:

- At a fixed dose, the incidence depends on coverage – for instance, the dose of 0.1 (dashed line in Fig. 5.3) corresponds to about 0.02% incidence at 90% coverage, 0.1% incidence at 95% coverage and more than 1% incidence at 99% coverage.

- The degree of uncertainty (distance between the LCLs and UCLs in the figure) increases as  $I$  decreases. This is caused by the fact that, as  $I$  decreases, the LCL (conservative bound) of the  $HD_M^I$  decreases more rapidly than the UCL (anti-conservative bound).



**Fig. 5.3: APROBA screenshot showing the relationship between  $HD_M^I$  and incidence  $I$ , for different levels of coverage.** The scenario for a subchronic NOAEL in a rat, for a deterministic quantal end-point, is illustrated here. The curves reflect the confidence bounds at various levels of coverage; for example, the leftmost (coverage 99%) and rightmost (coverage 1%) curves reflect a confidence interval with (two-sided) confidence of 98%.

- At a fixed coverage, the incidence  $I$  changes rapidly with dose. For instance, at 95% coverage, a change in  $HD_M^I$  from 0.1 to 1 leads to a change in  $I$  from 0.1% to 10%. In other words, increasing a probabilistic RfD (at 95% coverage) from 0.1 to 1 dose units would be associated with a decrease in the protected fraction of the population from 99.9% to 90%.
- At a fixed incidence, the  $HD_M^I$  for 99% coverage is less than an order of magnitude lower than the  $HD_M^I$  for 90% coverage.

Because the APROBA tool takes the PoD at a fixed  $M$  as an input, it cannot depict the impact of changing  $M$ . This would require deriving multiple PoDs corresponding to different values of  $M$  and could be accomplished in a full probabilistic approach. However, more advanced visualization techniques would be needed to simultaneously display the effect of  $M$  along with the effects of  $I$  and coverage.

## 6. INTERPRETATION AND USE OF RESULTS

The results of hazard assessment should include information on uncertainty, because it is needed by risk managers. As stated by Codex (2013):

Precaution is an inherent element of risk analysis. Many sources of uncertainty exist in the process of risk assessment and risk management .... The degree of uncertainty and variability in the available scientific information should be explicitly considered in the risk analysis.

However, these are not the only considerations relevant for decision-making. Codex (2013) defines risk management as:

The process, distinct from risk assessment, of weighing policy alternatives, in consultation with all interested parties, considering risk assessment and other factors relevant for the health protection of consumers and for the promotion of fair trade practices, and, if needed, selecting appropriate prevention and control options.

The European Union Food Regulation (EC) No. 178/2002 defines risk management in a similar way, as “the process ... of weighing policy alternatives ... considering risk assessment and other legitimate factors”, and states in its preamble that legitimate factors may include societal, economic, traditional, ethical and environmental factors and the feasibility of controls (EC, 2002). Similarly, the USEPA (2000) states that most risk management decisions are informed by a variety of factors, including scientific factors (risk assessment), economic factors, public values, political factors, social factors and technological factors. Thus, the results of hazard assessment need to be interpreted in relation to a wide range of other factors relevant to decision-making, which will vary depending on the regulatory context.

Currently, the protection goals are only occasionally specified. The incidence (extra risk) is often specified for genotoxic carcinogens (e.g.  $10^{-5}$  or  $10^{-6}$  for the general population), but not for other compounds or effects. When the BMD approach is used, the typical coverage used is 95%, but this coverage relates only to the BMDL, not to the health-based guidance value. Further, when the BMD approach is used, certain conventions or recommendations may exist as to which  $M$  would be considered as the critical or benchmark response (BMR) (see EFSA, 2009; USEPA, 2012). However, for non-cancer end-points in particular, the desired protection goals in terms of incidence  $I$  and magnitude  $M$ , together with the required coverage, are hardly ever specified in an explicit way, even in elaborate chemical risk management programmes.

It is noted that for the sake of transparency as well as comparability of risk assessments, both within the same and between different regulatory programmes, discussion on these issues appears highly desirable.

As the approaches described in this document are designed to make the protection goal and uncertainty explicit, they provide an improved basis for the weighing of hazards and their uncertainty against other factors relevant for risk management decision-making. The principal improvements are as follows:

- The protection goal is characterized explicitly in terms of the magnitude (severity) of the hazard and its incidence (proportion of target population affected) at the RfD.

- The degree of uncertainty in estimating the RfD (how much higher it might be) and its coverage (the probability that it provides the intended level of protection) is quantified.
- In decision-making, results can be readily calculated for alternative choices of the protection goal. This enables the risk manager to balance particular choices of the protection goal and of the uncertainties involved against socioeconomic interests. It also makes communication with stakeholders more transparent.
- Information can be included on other uncertainties, which are not or cannot be quantified, so that they can be taken into account when interpreting the results.

These advances clearly distinguish the estimation of the  $HD_M^I$  and its uncertainty (which are scientific considerations) from the setting of the protection goal ( $M$  and  $I$ ) and decisions on which coverage and degree of uncertainty are acceptable (which are risk management considerations, although toxicologists may need to assist in interpreting  $M$ ). This will help to ensure that the scientific basis of risk analysis is not affected by the consideration of other factors that are relevant for risk management, thus respecting the appropriate separation between risk assessment and risk management (see Codex, 2013: p. 108).

The following sections illustrate how the results of hazard assessment can be interpreted and how they might be weighed against other factors relevant for decision-making. This is done for several different contexts: setting a health-based guidance value, assessing the risk for a given situation and prioritization of chemical risks.

## 6.1 Setting a health-based guidance value

As a first example, consider a hazard characterization that aims to derive a health-based guidance value, such as an RfD (including chronic or acute), an ADI or a TDI. For the remainder of this section, this type of hazard characterization is illustrated in terms of an RfD. The traditional approach to deriving such values is by dividing a point of departure (e.g. a NOAEL or BMDL) by a composite uncertainty factor to determine a level that can be ingested over a defined time period (e.g. lifetime or 24 hours) without appreciable health risk (FAO/WHO, 2009). This deterministic approach results in a single value for the RfD. For a deterministic RfD, the coverage and degree of uncertainty (see sections 2 and 3) cannot be quantified, and usually the protection goals in terms of magnitude of effect ( $M$ ) and population incidence ( $I$ ) are specified only in a general way (e.g. dose without appreciable risk over lifetime for the general population, including sensitive subpopulations), but not quantified explicitly.

In contrast, a probabilistic hazard characterization provides an uncertainty distribution for the  $HD_M^I$  – that is, the human dose associated with a fraction  $I$  of the population being subject to an effect of magnitude or severity  $M$ . This distribution may be used to calculate the coverage of the deterministic RfD and the degree of uncertainty associated with it. It may also be used to derive a probabilistic RfD in which the protection goals in terms of coverage,  $M$  and  $I$  are explicitly specified. For the present example, where it is assumed that such protection goals have been predefined and transparently documented, the following situations may arise:

- If the coverage of the deterministic RfD is close to the predefined value (while taking account of any unquantified uncertainties), the deterministic RfD may be used for risk characterization.

- In cases where the coverage of the deterministic RfD is lower than the predefined value, a probabilistic RfD might be set by selecting a lower percentile of the  $HD_M^I$  distribution corresponding to the (higher) coverage considered acceptable and used for risk characterization.
- If the coverage of the deterministic RfD is found to be higher than the predefined value, then this value may be used for risk characterization. Alternatively, one might use the probabilistic RfD (at the predefined coverage), while taking account of any unquantified uncertainties, if applicable.
- If the degree of uncertainty is high (e.g. more than 100-fold, but depending on case-specific considerations), performance of a higher-tier hazard characterization might be considered, to avoid a health-based guidance value that is much lower than might be needed for achieving the protection goals. In this case, any knowledge or expectation of human exposure might be taken into account. For instance, if human exposure is expected to be orders of magnitude lower than the so far derived deterministic or probabilistic RfD, there is no need to use resources to obtain an RfD with a smaller degree of uncertainty.

Below, these considerations are further explained based on an excerpt from the case-study on deoxynivalenol, provided in Annex 5.

Consider first the tier 1 evaluation of the effect on body weight as per section A5.2 of the deoxynivalenol case-study. In this example, the predefined protection goals were as follows: critical magnitude of effect  $M = 5\%$ , critical incidence in the target human population  $I = 1\%$ , desired level of coverage = 95%. In other words, the  $HD_M^I$  is defined as  $HD_{05}^{01}$ , with 95% coverage.

The deterministic RfD of  $0.4 \mu\text{g}/\text{kg bw}$  per day for the critical effect was derived from a LOAEL extrapolated to a NOAEL (by an additional assessment factor of 3) and using default assessment factors of 10 for both interspecies and intraspecies differences.

Subsequently, an approximate probabilistic uncertainty analysis was performed using the APROBA tool, the results of which are given in Fig. 6.1.

The deterministic RfD was associated with a coverage of only about 85%. Thus, its coverage was clearly lower than the desired value of 95%. This case therefore corresponds to the second bullet above; consequently, the probabilistic RfD at the desired coverage level (95%) was found to be lower than the deterministic one – that is,  $0.15 \mu\text{g}/\text{kg bw}$  per day.

However, the degree of uncertainty was calculated to be about 200; in other words, the distance between the lower-bound estimate of the  $HD_{05}^{01}$  (the probabilistic RfD) and the upper-bound estimate was about 200-fold. This large uncertainty implies that a refined hazard characterization has the potential to reduce the overall uncertainty, resulting in a narrower confidence interval around the  $HD_{05}^{01}$  (and also perhaps a higher lower-bound estimate/probabilistic RfD). In particular, the results of the (approximate) probabilistic analysis suggested that a refined tier 2 assessment should be based on the BMD instead of the LOAEL/NOAEL, because the uncertainty in the associated hazard characterization aspects (using a NOAEL that was, in addition, extrapolated from a LOAEL) was estimated to have made the largest contribution to overall uncertainty. For more details, see the full text of the deoxynivalenol case-study in Annex 5.

NON-PROBABILISTIC ANALYSIS OUTPUTS <sup>i,k</sup>			
Target Human Dose (HD <sub>M</sub> <sup>l</sup> )	LCL	0.00899	µg/kg bw per day
	UCL	532	µg/kg bw per day
Fold Range of Uncertainty		59187	
Estimated "Coverage" of Non-Prob. LCL of HD <sub>M</sub> <sup>l</sup> *			100.0%
*Based on approximate probabilistic analysis, below.			
APPROXIMATE PROBABILISTIC ANALYSIS OUTPUTS			
Standard Confidence Interval			
Target Human Dose (HD <sub>M</sub> <sup>l</sup> )	LCL (P05)	0.154	µg/kg bw per day
	UCL (P95)	31.0	µg/kg bw per day
Degree of Uncertainty (Fold Range)			201
Estimated "Coverage" of Deterministic RfD			85%
Probabilistic RfD	= Approximate probabilistic HD <sub>M</sub> <sup>l</sup> at specified % confidence		
0.15	= Estimate of dose (µg/kg bw per day) at which, with		
	95%	confidence	
	1%	of the population will have	Example calculation
	of magnitude	≥	5%

Fig. 6.1: APROBA probabilistic results related to a deterministic assessment of deoxynivalenol (DON) using the NOAEL approach, with reduction of body weight compared with controls as the critical end-point. For details, see the DON case-study, provided in Annex 5.

Table 6.1 provides an overview of what was gained by going from a tier 1 to a tier 2 analysis in this example. As the fourth column shows, going to a higher-tier analysis did indeed greatly reduce the degree of uncertainty associated with the HD<sub>05</sub><sup>01</sup> (i.e. the width of its 90% confidence interval), by about 5-fold (from about 200 to about 40). Considering all of the available dose–response information for the critical effect by using the BMD approach, the higher-tier assessment resulted in a higher probabilistic RfD at the desired coverage level of 95% (0.44 instead of 0.15 µg/kg bw per day).

Table 6.1: Summary of uncertainty evaluation of BMDL-based assessments

Tier	Deterministic RfD (µg/kg bw per day)	Coverage (%)	Degree of uncertainty	Probabilistic RfD <sup>a</sup> (µg/kg bw per day)
1	0.4	85	200	0.15
2	1.7	68	43	0.44

<sup>a</sup> For incidence I = 1%, M = 5% decrease in body weight and coverage = 95%.

It is important to consider uncertainties that are usually not (explicitly) covered in the overall assessment factor (including secondary uncertainties that might be relevant for the specific case). This is illustrated in the full version of the deoxynivalenol case-study, including for further types of end-points and hazard characterization aspects.

## 6.2 Acceptability and communication of health risks for a given situation

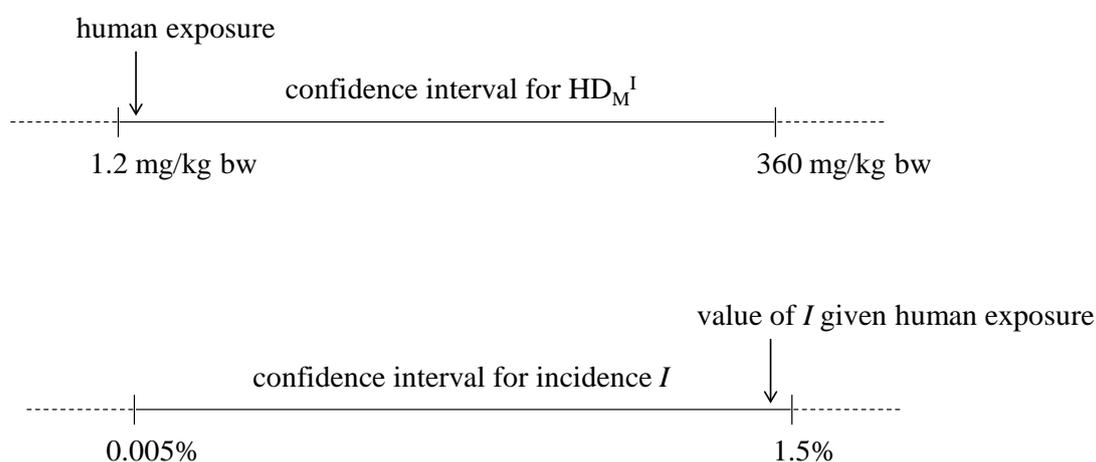
Given a measured, modelled or assumed exposure level, a decision-maker may want to know what type, magnitude and incidence of health effects might reasonably be expected in the exposed population given current conditions. Further, if decision-makers consider that the

estimated health effects are unacceptably large, they may want to know what health effects in the population might reasonably be expected after a particular action.

The purpose of the risk assessment at a given exposure may be 2-fold: (1) to inform decisions by a regulatory agency and (2) to provide regulatory officials with information for communicating with the public about current risks and the expected risks after taking a particular action.

In situations like this, the simplest form of assessment is to compare the estimated exposure with a deterministic or probabilistic health-based guidance value. If the health-based guidance value refers to acceptable levels of  $M$  and  $I$ , while the coverage is appropriate, and if the margin between the health-based guidance value and the exposure is sufficient to cover any uncertainties that have not been quantified, then the risk may be considered acceptable.

If these conditions are not met, then the risk manager may want to better understand what the exceedance of the  $HD_M^I$  means. Here, two possible approaches are illustrated. One approach is to compare the human exposure level with the whole confidence interval  $HD_M^I$ , as illustrated in the upper part of Fig. 6.2.



**Fig. 6.2: Communicating the potential health risk given a current exposure.** In the upper panel, the exposure is compared with a confidence interval for the  $HD_M^I$ , indicating that exposure could very well be much lower than the  $HD_M^I$ , even though it might be higher than the lower-bound estimate. In the lower panel, the value of  $I$  according to the pre-specified protection goal (in this case, 1%) is compared with the confidence interval for the estimated  $I$  at the current exposure, indicating that the incidence could very well be much lower than the pre-specified value of 1%, and, if not, exceed it only to a limited extent.

In this example, the risk manager might argue that although human exposure exceeds the lower bound of the confidence interval (and hence the probabilistic health-based guidance value, in line with what was already established), it is made visible that it is quite likely that the  $HD_M^I$  is much higher than human exposure. Another option that the risk manager might take is to consider the confidence interval for the incidence (given a specified  $M$ ) associated with the human exposure level. This is illustrated in the lower part of Fig. 6.2, where it can be seen that the incidence at the given exposure level could very well be much lower than the maximum accepted incidence (in this case, specified as 1%), while at the same time the risk manager can see that the worst-case estimate is 1.5% – that is, somewhat larger than 1% (at the confidence level used). In situations where other interests are involved, the risk manager might decide that the protection goals will at least be closely realized, while there is a good

chance that they will be realized amply. This consideration may be balanced with other societal interests.

In cases where human exposure turns out to be more in the middle of the  $HD_M^I$  confidence range, while the degree of uncertainty is high, the risk manager might decide that the uncertainties need to be reduced before taking action.

Basically, the approach of doing a refined assessment in case of large uncertainties was also followed in the deoxynivalenol case-study, as already summarized above, where for the end-point body weight, exposure of the target human population was estimated to be 0.44  $\mu\text{g}/\text{kg}$  bw per day, which was higher than the probabilistic RfD from the first-tier assessment, implying that the protection goals were not achieved. The second-tier assessment resulted in a considerably lower degree of uncertainty in the  $HD_M^I$ , whereas the probabilistic RfD was found to be equal to the human exposure of 0.440  $\mu\text{g}/\text{kg}$  bw per day. Thus, after the refined assessment, it was found that the protection goal was actually achieved (although without any further margin of safety). Of course, in all the examples discussed, the impact of additional uncertainties would need to be examined in order to conclude as to whether the estimated risk was considered acceptable.

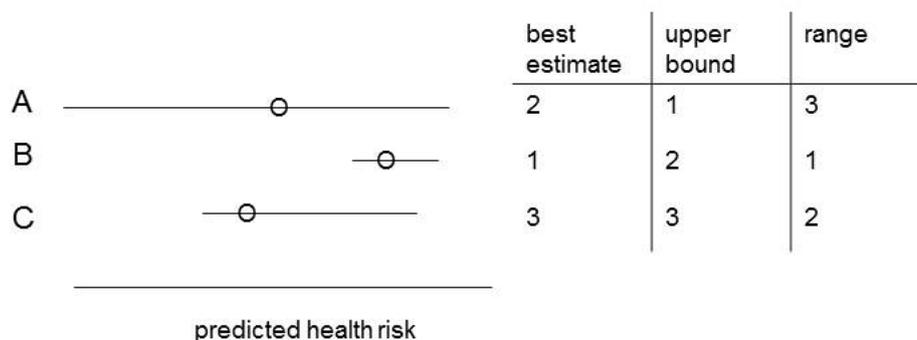
In the above discussion of the illustrative examples, exposure was assumed to be a single number, which might be an estimate of a UCL relating to the median individual or to a percentile of the population. Obviously, the best approach would be to take the uncertainties and variability in exposure into account as well. This may be done by conducting a full IPRA that quantifies variability and uncertainty in exposure as well as hazard. The final outcome of this could be an uncertainty distribution for the incidence (risk) in the population, for a specified type (and magnitude) of effect. This uncertainty distribution could be reduced to a confidence interval or uncertainty range of potential risks accounting for the sources of uncertainty and variability in hazard and exposure aspects simultaneously and in a similar way. As another advantage, performing a full IPRA makes visible which aspects of the risk assessment are most uncertain. It might well be that reducing uncertainties in exposure would be much more effective (and efficient) than generating additional toxicity data. With a full picture of the relative contributions from all sources of uncertainty, a rational choice can be made as to how best to use the available resources. For an illustration of a full IPRA applied to various substances (including deoxynivalenol), see Bokkers et al. (2009).

Procedures that provide decision-makers with only an upper confidence limit on the risk will leave them no choice other than to communicate a worst-case risk estimate to the public, without being able to show what this “worst case” means. In addition, single risk numbers, even when communicated to be worst-case estimates only, might easily be taken as representing the “true” risk by the general public. When provided with both an upper and lower risk estimate, the decision-maker can articulate that the risk might be relatively high, with some small probability, but that it is more likely to be much lower and (if applicable) that there is a fair probability that it is extremely low. The latter would not be visible, or understood, when only the upper-bound risk was communicated.

### 6.3 Prioritization of chemicals based on predicted risks

Usually, there are only limited resources for measures aiming at mitigating potential health effects from exposure to chemicals. Therefore, it may be useful to rank different chemicals, based on the available information on potential health risks. This information may, next to

other considerations, be used in prioritizing chemicals. As risk estimates are uncertain, ranking based on single (deterministic) values can easily lead to inadequate ranking of the real risks.



**Fig. 6.3: Prioritization based on best estimate, conservative (upper bound) or whole uncertainty range**

This is illustrated in Fig. 6.3, where, for three chemicals (A, B, C), the “best” estimate of the risk is indicated by a circle, together with an uncertainty range.

On the right-hand side, the ranking is shown when based on the best estimate, on the upper bound or on the whole range, respectively. Comparing chemicals A and B, the upper-bound risk estimate would rank A over B, but the best estimate would rank B over A. The appropriate way of ranking would be, however, to consider both the lower- and upper-bound risk estimates, indicating that the risk in A could be quite low, while such is not the case in B. Therefore, from a health perspective, investment in mitigating measures is more likely to be beneficial for B than for A. Comparing chemicals A and C, the upper-bound risk estimate and the best estimate would rank A over C. However, considering the full uncertainty range, one might rank C over A, based on the same argument as just discussed for A and B.

## 6.4 Socioeconomic analysis

Another approach to recognizing limited governmental or societal resources for reduction of chemical risks (or even for other public goals) is to evaluate the net societal benefits of taking action. In fact, under frameworks such as the European Union’s regulation on the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), socioeconomic analysis is a fully integrated part of the risk assessment of substances of concern. Similarly, in the USA, estimates of net economic benefits are often required as part of regulatory impact analyses supporting a major regulation.

Although it would be beyond the scope of this document to delve more deeply into the concepts behind socioeconomic analysis (which, at present, are also still evolving), its basic aim is to weigh the costs (in terms of necessary risk mitigation measures or of removing a substance from the market that offers societal benefits, such as mobility or availability of information technology) against the health benefits from regulating the chemical under question.

The underlying paradigm of this procedure, where health benefits from regulating a chemical are monetized and compared with the cost of technical investments in risk mitigation

measures, is again outside the scope of this document. Nevertheless, compared with deterministic hazard characterization approaches, the approach for characterizing uncertainty as presented in this document provides a much more suitable framework upon which socioeconomic analyses can be based. Specifically, the  $HD_M^I$  provides information on the magnitude  $M$  and incidence  $I$  of effect that, in principle, can be monetized much more readily than a PoD or a deterministic health-based guidance value. Moreover, because  $M$  and  $I$  are made explicit, the net benefits of different risk management options, with different costs and leading to different values of  $M$  and  $I$ , could be compared (e.g. to identify the option with the greatest net benefits).

## 7. NEXT STEPS

This section proposes next steps with regard to further research on methodological issues and data input and with regard to implementation and communication of the method described in this guidance.

### 7.1 Further research

#### **7.1.1 Methodological issues**

There are a number of aspects for which the associated uncertainty distribution has not been assessed so far, such as for missing studies, read-across, kinetic aspects in route-to-route extrapolation, in silico, in vitro and epidemiological approaches, etc. Specific methodology will probably need to be developed to quantify the uncertainties associated with such types of approaches in hazard characterization. The assessment of uncertainty in the hazard characterization for mixtures and for cumulative exposure to multiple stressors will also need to be investigated further.

In addition, research is needed to resolve the methodological problems that are shared by most of the research efforts aiming to quantify uncertainties. One of the common problems is that the observed variance of equipotent doses among chemicals is the sum of the real variance plus the measurement/estimation variance. Work is needed to develop the methodology that corrects the observed variance in equipotent dose (BMD) ratios for the size of the measurement/estimation variance in the underlying estimates of equipotent doses (BMDs). If such methodology were available, it could be applied to the examination of uncertainties for any of the aspects, resulting in more accurate estimates of the real underlying chemical-to-chemical variation. For instance, Bokkers & Slob (2007) corrected the observed variation in BMD ratios by the mean of the estimated standard errors of the individual BMDs and in this way achieved a lower GSD for the variation among chemicals. This is a rough approach only and should be validated and, possibly, improved. Another, more general issue is that of the representativeness of the chemicals informing the uncertainty distribution for a given aspect. Is it possible to define categories of chemicals, grouped by chemical attributes, mode of action, metabolic pathway or end-points, or on some other basis for which it is reasonable and possible to develop tailor-made probabilistic assessment factors? This would increase confidence that a specific chemical at hand (lacking information for that aspect) can reasonably be treated as a random draw from that population of chemicals. This question might be answered in the future by analysis of the existing (or expanded) database of observations for each aspect and discriminating groups of chemicals. If distributions are derived for an attribute such as common chemical structure, a methodological issue will be whether the distribution is sufficiently different from a broader classification to be an improvement. If distributions are derived for groups of chemicals based on an attribute on which scientists may have differing opinions, such as mode of action, when data are limited, a methodological question will be how to reflect uncertainty as to whether the chemical is a member of that category or whether a less targeted distribution should be used.

Another issue applies to the representativeness of human subjects who are studied and who serve as a basis for estimating intraspecies variation, compared with the exposed population. In general, the data underlying the estimation of intraspecies variation relate to healthy volunteers, a subpopulation that will most likely show smaller variation than the overall

human population. What adjustments should be made to the observed intraspecies variability to reasonably represent the full diversity of the human population?

The approximate probabilistic analysis is likely to be more frequently used by risk assessors than the full probabilistic approach. It may also be that for some time, even the approximate probabilistic analysis may be done as a companion to a traditional non-probabilistic approach. Research that systematically compares the outcomes of these approaches for a range of scenarios would provide insight into the extent to which the approximate probabilistic analysis might deviate from the full probabilistic approach and how both compare with traditional approaches (e.g. in terms of the resulting  $HD_M^1$ ).

The issue of secondary uncertainties has been discussed to a limited degree in this document. Quantitative studies on secondary uncertainty may reveal to what extent the outcomes from a probabilistic hazard characterization are sensitive to various assumed uncertainty distributions for each single aspect. In addition, such analyses may evaluate the combined effect of secondary uncertainties. Thus, insight may be obtained into the “overall” secondary uncertainty, as well as the relative contribution of each aspect to the overall secondary uncertainty.

The measurement and identification of uncertainties in toxicological research and transfer of such measurements into risk assessment should be encouraged, through the WHO Chemical Risk Assessment Network and other groups.

Development of Bayesian methods may be of use because of their ability to incorporate the uncertainty in prior assumptions that enable one to integrate primary and secondary uncertainties or combine uncertainties from chemical-specific data with the generic uncertainties derived from data on many chemicals for the same aspect.

### **7.1.2 Input data**

The approach of evaluating individual uncertainties and combining them into the overall uncertainty of the final outcome of a hazard characterization hinges on the validity of the individual uncertainties for each aspect. Section 4 evaluated the individual uncertainties for various aspects based on available reviews of historical data. The currently available reviews did not always directly aim to evaluate the uncertainty of the relevant aspect in the context of a probabilistic hazard characterization, and therefore the results were not always fully fit for purpose. If the significance of the approach of evaluating uncertainties as laid out in this document is recognized, the need for more and better reviews of historical data becomes evident. Therefore, the collection of relevant historical data and the analysis of PoD ratios relating to any of the aspects potentially involved in hazard characterizations would be a useful endeavour.

A number of priorities may be formulated for research on the following issues, considered as highly relevant and relatively feasible, in no particular order:

- *Oral NOAEL to BMDL uncertainty*: Examine variation in oral NOAEL to BMDL ratios for various study types (among others, for quantal end-points in repeated-dose studies and for continuous end-points in subacute studies);
- *Generic BMD uncertainty (second step of NOAEL uncertainty, see section 4.2.2)*: Examine the variation in BMD uncertainty (BMDU/BMDL) for various study types and for both response types;

- *Inhalation NOAEL to BMDL uncertainty*: Examine NOAEL to BMDL ratios for various study types and for both response types;
- *Interspecies uncertainty for inhalation*: Examine variation in inhalation BMD ratios for pairs of species;
- *Missing study uncertainty*: Examine the variation in BMD ratios between two study types (e.g. developmental versus subchronic); and
- *Intraspecies uncertainty*: Expand the database of relevant human data, and further examine the across-chemical variation in interindividual variability.

Different distributions for the same aspect have been developed worldwide, and this is likely to continue. Systematic review for database expansion and harmonization of such distributions as well as identification and filling of gaps in knowledge should be encouraged using, among others, the WHO Chemical Risk Assessment Network. A portal with international stature should be developed to collect, update and provide access to data used for distribution development and to example applications.

Any studies aimed at improving or establishing generic distributions should, to the extent possible, include evaluation of secondary uncertainty associated with the proposed distribution so that this can be taken into account when it is used in risk assessment.

Pending additional research into the above issues, it may be useful to explore some of these using expert elicitation. Such an approach cannot necessarily substitute for explicitly data-driven research, but it can capture in a systematic way the nature and range of expert opinion on some of the issues, given current information and understanding. It might highlight areas in which there is currently relative consensus or lack thereof among experts (in contrast, consensus itself cannot be taken as a criterion, and care must be taken to include up-to-date science in the assessment and not “cement” established practice that “has always been done like this”).

## 7.2 Implementation and dissemination

Implementation and dissemination of the methodology in this guidance hinge on two aspects: communication issues and training issues.

Target groups are risk assessors and risk managers worldwide in, among others, regulatory authorities, industry or expert panels of international institutions, such as WHO/IPCS or the Organisation for Economic Co-operation and Development (OECD).

This document provides measures of uncertainty (coverage, degree of uncertainty) coupled to useful hazard/risk-related parameters, such as severity, incidence and fraction of the population protected, that may be used as tools for communicating uncertainty to the risk manager as well as to the general public. The method allows exploration of a variety of risk management options. In the examples, it was briefly indicated how this might work in specific cases. However, these are theoretical exercises, and it would be valuable to have more systematic studies on how risk communication of uncertainties, using the tools presented in this document, functions in practice, regarding both risk managers and other stakeholders, such as the general public. The development of case-studies is essential to further demonstrate that the approach is a useful tool for risk management, including socioeconomic analysis. Dissemination of this approach can be further favoured in various

ways – for example, through websites, brochures, training and training materials, conferences, workshops, the WHO Chemical Risk Assessment Network and scheduled webinars.

Whereas this document concentrates on uncertainty in hazard characterization, clearly this uncertainty should be connected to uncertainty in exposure to be able to arrive at risk values in terms of impacts that are meaningful to risk managers and other stakeholders and in risk communication. It is therefore recommended that the results of this IPCS project be integrated with those from other IPCS projects, notably the one on uncertainty in exposure assessment (IPCS, 2008).

## 8. REFERENCES

Allen BC, Kavlock RJ, Kimmel CA, Faustman EM (1994). Dose–response assessment for developmental toxicity. II. Comparison of generic benchmark dose estimates with no observed adverse effects levels. *Fundam Appl Toxicol.* 23:487–95.

Baird JS, Cohen JT, Graham JD, Shyakhter AI, Evans JS (1996). Noncancer risk assessment: probabilistic characterization of population threshold doses. *J Hum Ecol Risk Assess.* 2(1):79–102.

Bokkers BG, Slob W (2005). A comparison of ratio distributions based on the NOAEL and the benchmark approach for subchronic-to-chronic extrapolation. *Toxicol Sci.* 85(2):1033–40.

Bokkers BGH, Slob W (2007). Deriving a data-based interspecies assessment factor using the NOAEL and the benchmark dose approach. *Crit Rev Toxicol.* 37:353–77.

Bokkers BGH, Bakker MI, Boon PE, Bos P, Bosgra S, Van der Heijden GWAM et al. (2009). The practicability of the integrated probabilistic risk assessment (IPRA) approach for substances in food. Bilthoven: National Institute for Public Health and the Environment (RIVM report 320121001/2009; <http://www.rivm.nl/bibliotheek/rapporten/320121001.pdf>, accessed 18 November 2017).

Chiu WA (2011). Trichloroacetic acid: updated estimates of its bioavailability and its contribution to trichloroethylene-induced mouse hepatomegaly. *Toxicology.* 285(3):114–25.

Chiu WA, Slob W (2015). A unified probabilistic framework for dose-response assessment of human health effects. *Environ Health Perspect.* 123(12):1241–54.

Chiu WA, White P (2006). Steady-state solutions to PBPK models and their applications to risk assessment. I: Route-to-route extrapolation of volatile chemicals. *Risk Anal.* 26:769–80.

Chiu WA, Okino MS, Evans MV (2009). Characterizing uncertainty and population variability in the toxicokinetics of trichloroethylene and metabolites in mice, rats, and humans using an updated database, physiologically based pharmacokinetic (PBPK) model, and Bayesian approach. *Toxicol Appl Pharmacol.* 241(1):36–60.

Codex (2013). Working principles for risk analysis for application in the framework of Codex Alimentarius. In: Procedural manual, 21st edition. Rome: Food and Agriculture Organization of the United Nations and World Health Organization, Joint FAO/WHO Food Standards Programme, Codex Alimentarius Commission; 107–13 (<http://www.fao.org/3/a-i3243e.pdf>, accessed 18 November 2017).

EC (2002). Regulation (EC) No 178/2002 of the European Parliament and of the Council of 28 January 2002 laying down the general principles and requirements of food law, establishing the European Food Safety Authority and laying down procedures in matters of food safety. OJ (Off J Eur Communities). L 31:1–24 (<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32002R0178>, accessed 18 November 2017).

EC (2013). Making risk assessment more relevant for risk management. European Commission, Scientific Committee on Health and Environmental Risks, Scientific Committee on Emerging and Newly Identified Health Risks and Scientific Committee on Consumer Safety ([http://ec.europa.eu/health/scientific\\_committees/consumer\\_safety/docs/sccs\\_o\\_130.pdf](http://ec.europa.eu/health/scientific_committees/consumer_safety/docs/sccs_o_130.pdf), accessed 18 November 2017).

Edler L, Hart A, Greaves P, Carthew P, Coulet M, Boobis A et al. (2013). Selection of appropriate tumour data sets for benchmark dose modelling (BMD) and derivation of a margin of exposure (MoE) for substances that are genotoxic and carcinogenic: considerations of biological relevance of tumour type, data quality and uncertainty assessment. *Food Chem Toxicol.* 70:264–89.

EFSA (2009). Scientific opinion. Use of the benchmark dose approach in risk assessment. Guidance of the Scientific Committee (Question No EFSA-Q-2005-232). *EFSA J.* 1150:1–72 (<http://onlinelibrary.wiley.com/doi/10.2903/j.efsa.2009.1150/epdf>, accessed 18 November 2017).

EFSA (2014). Guidance on expert knowledge elicitation in food and feed safety risk assessment. *EFSA J.* 12 (6):3734. Parma: European Food Safety Authority (<http://onlinelibrary.wiley.com/doi/10.2903/j.efsa.2014.3734/epdf>, accessed 18 November 2017).

FAO/WHO (2009). Principles and methods for the risk assessment of chemicals in food. A joint publication of the Food and Agriculture Organization of the United Nations and the World Health Organization. Geneva: World Health Organization (Environmental Health Criteria 240; [http://www.inchem.org/documents/ehc/ehc/ehc240\\_index.htm](http://www.inchem.org/documents/ehc/ehc/ehc240_index.htm), accessed 18 November 2017).

Hart A, Gosling JP, Boobis A, Coggon D, Craig P, Jones D (2010). Development of a framework for evaluation and expression of uncertainties in hazard and risk assessment. Prepared by the Food and Environment Research Agency for the United Kingdom Food Standards Agency (FSA Project No. T01056; [https://www.food.gov.uk/sites/default/files/676-1-1148\\_T01056\\_Final\\_Report\\_for\\_Web.pdf](https://www.food.gov.uk/sites/default/files/676-1-1148_T01056_Final_Report_for_Web.pdf), accessed 18 November 2017).

Hattis D, Lynch MK (2007). Empirically observed distributions of pharmacokinetic and pharmacodynamic variability in humans – Implications for the derivation of single point component uncertainty factors providing equivalent protection as existing RfDs. In: Lipscomb JC, Ohanian EV, editors. *Toxicokinetics in risk assessment*. Informa Healthcare USA, Inc.:69–93.

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://apps.who.int/iris/bitstream/10665/43294/1/9241546786\\_eng.pdf](http://apps.who.int/iris/bitstream/10665/43294/1/9241546786_eng.pdf), accessed 18 November 2017).

IPCS (2008). Guidance document on characterizing and communicating uncertainty in exposure assessment. In: *Uncertainty and data quality in exposure assessment*. Geneva: World Health Organization, International Programme on Chemical Safety (Harmonization Project Document No. 6; [http://www.who.int/entity/ipcs/publications/methods/harmonization/exposure\\_assessment.pdf](http://www.who.int/entity/ipcs/publications/methods/harmonization/exposure_assessment.pdf), accessed 18 November 2017).

IPCS (2010). Principles of characterizing and applying physiologically based pharmacokinetic and toxicokinetic models in risk assessment. Geneva: World Health Organization, International Programme on Chemical Safety (Harmonization Project Document No. 9; [http://www.who.int/entity/ipcs/methods/harmonization/areas/pbpb\\_models.pdf](http://www.who.int/entity/ipcs/methods/harmonization/areas/pbpb_models.pdf), accessed 18 November 2017).

Janer G, Hakkert BC, Slob W, Vermeire T, Piersma AH (2007a). A retrospective analysis of the two-generation study: what is the added value of the second generation? *Reprod Toxicol.* 24:97–102.

Janer G, Hakkert BC, Piersma AH, Vermeire T, Slob W (2007b). A retrospective analysis of the added value of the rat two-generation reproductive toxicity study versus the rat subchronic toxicity study. *Reprod Toxicol.* 24(1):103–13.

Lehman AJ, Fitzhugh OG (1954). 100-fold margin of safety. *Assoc Food Drug Off US Q Bull.* 18(1):33–5.

Meek ME, Boobis A, Crofton MK, Heinemeyer G, Van Raaij M, Vickers C (2011). Risk assessment of combined exposure to multiple chemicals: a WHO/IPCS framework. *Regul Toxicol Pharmacol.* 60(2 Suppl):S1–S14.

Meek ME, Boobis A, Cote I, Dellarco V, Fotakis G, Munn S et al. (2013). New developments in the evolution and application of the WHO/IPCS framework on mode of action/species concordance analysis. *J Appl Toxicol.* 34(1):1–18.

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.

Schneider K, Schuhmacher-Wolz U, Hassauer M, Darschnik S, Elmshäuser E, Mosbach-Schulz O (2006). A probabilistic effect assessment model for hazardous substances at the workplace. *Regul Toxicol Pharmacol.* 44(2):172–81.

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.* 18:787–98.

Slob W, Bakker MI, Te Biesebeek JD, Bokkers BGH (2014). Exploring the uncertainties in cancer risk assessment using the integrated probabilistic risk assessment (IPRA) approach. *Risk Anal.* 34(8): 1401-1421.

USEPA (1994). Methods for derivation of inhalation reference concentrations (RfCs) and application of inhalation dosimetry. Washington (DC): United States Environmental Protection Agency, Office of Research and Development, Office of Health and Environmental Assessment (EPA/600/8-90/066F).

USEPA (1995). Policy for risk characterization. Washington (DC): United States Environmental Protection Agency (available as Appendix A in USEPA, 2000).

USEPA (1997). Guiding principles for Monte Carlo analysis. Washington (DC): United States Environmental Protection Agency, Risk Assessment Forum (EPA/630/R-97/001).

USEPA (2000). Risk characterization handbook. Washington (DC): United States Environmental Protection Agency, Science Policy Council (EPA 100-B-00-002; [https://www.epa.gov/sites/production/files/2015-10/documents/osp\\_risk\\_characterization\\_handbook\\_2000.pdf](https://www.epa.gov/sites/production/files/2015-10/documents/osp_risk_characterization_handbook_2000.pdf), accessed 18 November 2017).

USEPA (2001). Risk assessment guidance for Superfund: Volume III – Part A, Process for conducting probabilistic risk assessment. Washington (DC): United States Environmental Protection Agency, Office of Emergency and Remedial Response (EPA 540-R-02-002; [https://www.epa.gov/sites/production/files/2015-09/documents/rags3adt\\_complete.pdf](https://www.epa.gov/sites/production/files/2015-09/documents/rags3adt_complete.pdf), accessed 18 November 2017).

USEPA (2011). Recommended use of body weight<sup>3/4</sup> as the default method in derivation of the oral reference dose. Washington (DC): United States Environmental Protection Agency, Risk Assessment Forum (EPA/100/R11/0001; <https://www.epa.gov/sites/production/files/2013-09/documents/recommended-use-of-bw34.pdf>, accessed 18 November 2017).

USEPA (2012). Benchmark dose technical guidance. Washington (DC): United States Environmental Protection Agency, Risk Assessment Forum (EPA/100/R-12/001; [https://www.epa.gov/sites/production/files/2015-01/documents/benchmark\\_dose\\_guidance.pdf](https://www.epa.gov/sites/production/files/2015-01/documents/benchmark_dose_guidance.pdf), accessed 18 November 2017).

Van der Voet H, Slob W (2007). Integration of probabilistic exposure assessment and probabilistic hazard characterization. *Risk Anal.* 27:351–71.

Vermeire T, Stevenson H, Pieters MN, Rennen M, Slob W, Hakkert BC (1999). Assessment factors for human health risk assessment: a discussion paper. *Crit Rev Toxicol.* 29:439–90.

Vose D (2008). Risk analysis: a quantitative guide, 3rd edition. Chichester: John Wiley & Sons.

## 9. GLOSSARY OF TERMS<sup>15</sup>

**Acute exposure:** A contact between an agent and a target occurring over a short time, generally less than a day.

**Adjustment:** Extrapolating an equipotent dose from one situation to another for a given chemical. When there is no chemical-specific information, a default assessment factor is usually applied, which is assumed to be conservative; otherwise, a chemical-specific adjustment factor (CSAF) may be used.

**Aspect (of hazard characterization):** In this report, refers to the various parts of the hazard characterization with associated uncertainties, such as the point of departure, adjustments made due to characteristics of the study population or study design that differ from the target population or target conditions (e.g. interspecies differences, exposure duration), and the amount of variability due to heterogeneity in the human population.

**Assessment factor:** Numerical adjustment used to extrapolate from experimentally determined (dose–response) relationships to estimate the agent exposure below which an adverse effect is not likely to occur. In the probabilistic framework, the assessment factor is considered to be uncertain, and this uncertainty is reflected by an uncertainty distribution. Related terms: Safety factor, Uncertainty factor.

**Benchmark dose (BMD):** A dose of a substance associated with a specified (non-zero) effect, the benchmark response.

**Benchmark dose lower confidence limit (BMDL):** The lower bound of the confidence interval for the benchmark dose. The BMD confidence interval accounts for the uncertainty in the estimate of the dose–response due to limitations of the experimental design, such as limited sample size and limited number of doses. The BMDL can be used as the point of departure for derivation of a health-based guidance value or a margin of exposure.

**Benchmark dose upper confidence limit (BMDU):** The upper bound of the confidence interval for the benchmark dose.

**Benchmark response (BMR):** The specified non-zero effect defining the BMD. Examples include a 10% increase in incidence, a 5% decrease in red blood cells or mild liver lesions.

**Chemical-specific adjustment factor (CSAF):** A modification of the default 10-fold uncertainty factor, which incorporates appropriate data on species differences or human variability in either toxicokinetics (fate of the chemical in the body) or toxicodynamics (actions of the chemical on the body). In a probabilistic framework, the CSAF is an

---

<sup>15</sup> Adapted from the following sources: (1) this guidance document; (2) IPCS (2004). IPCS risk assessment terminology. Geneva: World Health Organization, International Programme on Chemical Safety (Harmonization Project Document No. 1); (3) IPCS (2008). Guidance document on characterizing and communicating uncertainty in exposure assessment. In: Uncertainty and data quality in exposure assessment. Geneva: World Health Organization, International Programme on Chemical Safety (Harmonization Project Document No. 6); and (4) FAO/WHO (2009). Principles and methods for the risk assessment of chemicals in food. A joint publication of the Food and Agriculture Organization of the United Nations and the World Health Organization. Geneva: World Health Organization (Environmental Health Criteria 240).

uncertainty factor, reflected by an uncertainty distribution accounting for uncertainties in the chemical-specific data.

**Chronic exposure:** A continuous or intermittent long-term contact between an agent and a target.

**Confidence interval:** An interval that is expected to enclose the true value of the parameter of interest with a specified confidence (e.g. 90%).

**Continuous response:** The value of a biological end-point observed (measured) in an individual subject. In theory, a continuous response can – as opposed to quantal end-points – take on any (positive) value within the biological limits of the end-point. In scientific practice, only quasi-continuous responses can be recorded, because of the natural limits of resolution of the measurement equipment.

**Coverage:** The per cent confidence that a given estimate of the target human dose (the  $HD_M^I$ ) is not lower than the “true” value of that dose.

**Degree of uncertainty:** The size of the uncertainty range (confidence interval) around an estimate of the target human dose ( $HD_M^I$ ), expressed as the ratio of the upper and the lower confidence limits of that estimate.

**Deterministic hazard characterization:** A hazard characterization where the calculations are based on single values (often conservative).

**Deterministic quantal data:** Data in the form of binary (yes/no) observations, conceptualized as resulting from a process in which an underlying continuous response has a cut-point, such that individuals whose underlying continuous response is above/below the cut-point are recorded as responding/non-responding.

**Distribution:** A probability distribution is a mathematical description of a function that relates probabilities to specified intervals of a continuous quantity, or to values of a discrete quantity, for a random variable. Probability distribution models can be non-parametric or parametric. Distributions such as normal, lognormal and others are examples of parametric probability distribution models, which can be fit to data sets by estimating their parameter values based upon the data.

**Dose–response assessment:** Analysis of the relationship between the total amount of an agent administered to a sample of biological units (e.g. subjects) and the changes developed in that sample of subjects in reaction to that agent.

**Effect metric:** A metric that quantifies the magnitude of change for an end-point compared with its background value. To enable estimation of an equipotent dose between populations or between individuals, the metric should be such that a given value can be considered to reflect a similar magnitude of effect even when the populations or individuals have different background values. An example of such an effect metric is the per cent change in a continuous end-point.

**Equipotent dose:** Dose that elicits the same magnitude  $M$  of the effect metric in different species or in different individuals of the same species.

**Exposure duration:** The length of time over which continuous or intermittent contacts occur between an agent and a target.

**Hazard characterization:** The qualitative and, wherever possible, quantitative description of the inherent property of an agent or situation having the potential to cause adverse effects. This should, where possible, include a dose–response assessment and its associated uncertainties. Hazard characterization is the second stage in the process of hazard assessment and the second of four steps in risk assessment.

**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. Hazard identification is the first stage in hazard assessment and the first of four steps in risk assessment.

**Health-based guidance value:** A numerical value derived by dividing a point of departure (a no-observed-adverse-effect level, benchmark dose or benchmark dose lower confidence limit) by a composite uncertainty factor to determine a level that can be taken up over a defined time period (e.g. lifetime or 24 hours) without appreciable health risk.

**Lowest-observed-adverse-effect level (LOAEL):** Lowest concentration or amount of a substance, found by experiment or observation, that causes an adverse alteration of morphology, functional capacity, growth, development or lifespan of the target organism distinguishable from normal (control) organisms of the same species and strain under the same defined conditions of exposure.<sup>16</sup>

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

**Margin of safety:** The margin between the health-based guidance value and the actual or estimated exposure dose or concentration.

**Mode of action:** A biologically plausible sequence of key events leading to an observed effect supported by robust experimental observations and mechanistic data. A mode of action describes key cytological and biochemical events – that is, those that are both measurable and necessary to the observed effect – in a logical framework.

**No-observed-adverse-effect level (NOAEL):** Greatest concentration or amount of a substance, found by experiment or observation, that causes no adverse alteration of morphology, functional capacity, growth, development or lifespan of the target organism distinguishable from those observed in normal (control) organisms of the same species and strain under the same defined conditions of exposure.<sup>17</sup>

---

<sup>16</sup> Note that although this is the official IPCS definition of the LOAEL (FAO/WHO, 2009), it needs to be understood that the phrase “Lowest concentration or amount” relates to the limited number of tested concentrations available in that particular experiment or observation and the phrase “that causes an adverse alteration” references alterations that would be observable with the specific study design and size.

<sup>17</sup> Note that although this is the official IPCS definition of the NOAEL (FAO/WHO, 2009), it needs to be understood that the phrase “Greatest concentration or amount” relates to the limited number of tested concentrations available in that particular experiment or observation. Furthermore, the phrase “that causes no adverse alteration ... distinguishable from ...” does not imply that effects are absent. As discussed in this document, the NOAEL has been shown to relate to effects in the order of 5% or 10%, on average over studies, whereas effects will be even greater in individual cases.

**Point of departure (PoD):** A value for the dose (obtained from relevant dose–response data) that serves as the starting point for estimating the equipotent dose in a human target population.

**Probabilistic:** An approach based on probability distributions of values rather than single point values. Related term: Deterministic.

**Probabilistic reference dose:** An estimate of the daily exposure dose where, with a given coverage (confidence), a fraction  $I$  of the human population is subject to an effect of magnitude  $M$  or greater during a lifetime.

**Problem formulation:** A process that describes the safety problem and its context, in order to identify those elements of hazard or risk associated with a chemical that are relevant to potential risk management decisions.

**Quantal response:** A change in a biological end-point in response to, for example, exposure to a chemical that – as opposed to a continuous response – can take on only two values in an individual (yes/no effect). Quantal responses are often deduced from observed responses that were originally scored in terms of severity categories, such as minimal, mild, moderate and severe (e.g. less than mild versus mild or more severe). Usually, quantal data are reported as the number of subjects with the effect out of the total number of subjects in the treatment group.

**Reference dose (RfD):** An estimate of the daily exposure dose that is likely to be without deleterious effect even if continued exposure occurs over a lifetime.

**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. The risk assessment process includes four steps: hazard identification, hazard characterization, exposure assessment and risk characterization. It is the first component in a risk analysis process.

**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. Risk characterization is the fourth step in the risk assessment process.

**Risk management:** Decision-making process involving considerations of political, social, economic and technical factors with relevant risk assessment information relating to a hazard so as to develop, analyse and compare regulatory and non-regulatory options and to select and implement appropriate regulatory response to that hazard. Risk management comprises three elements: risk evaluation, emission and exposure control, and risk monitoring.

**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. Related terms: Assessment factor, Uncertainty factor.

**Secondary uncertainty:** In this report, used to refer to uncertainty about the distributions used to quantify uncertainty.

**Sensitivity analysis:** A study of how the variation in the outputs of a calculation can be attributed to, qualitatively or quantitatively, different sources of variation in calculation inputs.

**Stochastic quantal data:** Data in the form of binary (yes/no) observations, conceptualized as resulting from a process in which each individual has a specific (individual) probability of showing the response or not, given the specific circumstances (including the dose of the chemical being studied).

**Subacute exposure:** A repeated contact between an agent for a duration between acute and subchronic exposure, usually up to 4 weeks.

**Subchronic exposure:** A contact between an agent and a target of intermediate duration between subacute and chronic, usually between a month and a year (depending on the species tested).

**Target human dose ( $HD_M^I$ ):** The human dose where a fraction  $I$  of the population shows an effect of magnitude (or severity)  $M$  or greater (for the critical effect considered).

**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. The term has essentially the same meaning as pharmacokinetics, but the latter term should be restricted to the study of pharmaceutical substances.

**Uncertainty:** Uncertainty in risk assessment in the general sense is defined by IPCS (2004) as “imperfect knowledge concerning the present or future state of an organism, system, or (sub)population under consideration”. In relation to the specific topic of this monograph, it can be further defined as lack of knowledge regarding the “true” value of a quantity, lack of knowledge regarding which of several alternative model representations best describes a system of interest, or lack of knowledge regarding which probability distribution function and its specification should represent a quantity of interest.

**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. Related terms: Assessment factor, Safety factor.

**Variability:** Heterogeneity of values over time, space or different members of a population, including stochastic variability and controllable variability. Variability implies real differences among members of that population. For example, different individual persons have different intake and susceptibility. In relation to human exposure assessment, differences over time for a given individual are referred to as intraindividual variability; differences over members of a population at a given time are referred to as interindividual variability.

## ANNEX 1: COMBINING UNCERTAINTIES FOR THE THREE APPROACHES

### A1.1 General approach to combining uncertainties

By definition, the  $HD_M^I$  associated with a given  $BMD_M$  is equal to:

$$HD_M^I = \frac{BMD_M}{\text{Aspect}_1 \cdot \text{Aspect}_2 \cdot \dots \cdot \text{Aspect}_n} \quad (\text{A1-1})$$

Here, the  $BMD_M$  refers to the “true value” of the animal dose resulting in effect  $M$  for a given critical end-point, and not an estimate of this value, such as the maximum likelihood estimate. Similarly, each  $\text{Aspect}_{1...n}$  denotes the true value of the adjustment for the relevant extrapolation (e.g. those listed in Table 3.1) that would be needed to convert the  $BMD_M$  into the (true value of the) target human dose,  $HD_M^I$ . In the probabilistic approach, each component in the right-hand side of equation A1-1 is reflected by an uncertainty distribution, resulting in an uncertainty distribution for the  $HD_M^I$ . This is analogous to the procedure used for deriving a deterministic health-based guidance value, such as an RfD, ADI or TDI, except that:

- the parameters in equation A1-1 are replaced by uncertainty distributions instead of conservative values;
- the magnitude of effect  $M$  is made explicit; and
- intraspecies variability depends explicitly on the target population incidence  $I$ .

For instance, for a chronic BMDL, the traditional approach would be:

$$\text{RfD} = \frac{\text{BMDL}}{AF_{\text{Inter}} \cdot AF_{\text{Intra}}} \quad (\text{A1-2})$$

Here, BMDL,  $AF_{\text{Inter}}$  (the assessment factor for interspecies extrapolation) and  $AF_{\text{Intra}}$  (the assessment factor for intraspecies variability) are all single (conservative) numbers.

In the probabilistic approach, the analogous equation to equation A1-2 would be:

$$HD_M^I = \frac{BMD_M}{AF_{\text{Inter-BS}} \cdot AF_{\text{Inter-TK/TD}} \cdot AF_{\text{Intra-I}}} \quad (\text{A1-3})$$

where each component is reflected by an uncertainty distribution. Note that the component AF in equation A1-3 relates to the unknown (uncertain) factor, whereas in equation A1-2 it relates to a conservative percentile of that uncertainty distribution. Further, the factor  $AF_{\text{Inter}}$  is divided into two subfactors,  $AF_{\text{Inter-BS}}$  (body size) and  $AF_{\text{Inter-TK/TD}}$  (remaining TK/TD differences), whereas  $AF_{\text{Intra-I}}$  indicates that intraspecies variability depends on the target population incidence  $I$ .

Similarly, for a NOAEL, the traditional approach would be:

$$RfD = \frac{NOAEL}{AF_{Inter} \cdot AF_{Intra}} \quad (A1-4)$$

Again, NOAEL,  $AF_{Inter}$  and  $AF_{Intra}$  are all single numbers.

When characterizing uncertainties, the underlying relationship is:

$$HD_M^I = \frac{NOAEL}{AF_{PoD-NOAEL} \cdot AF_{Inter-BS} \cdot AF_{Inter-TK/TD} \cdot AF_{Intra-I}} \quad (A1-5)$$

In this case, the NOAEL is still a single number, but an additional (uncertain) factor ( $AF_{PoD-NOAEL}$ ) needs to be included to account for the fact that the NOAEL is an uncertain estimate of the BMDL (see section 4.2.2 of the main document). Moreover, as noted in section 4.2.2, the uncertainty distribution for the additional aspect  $AF_{PoD-NOAEL}$  is anchored to a specific magnitude of effect  $M$ , which is (by assumption) a 5% relative change for continuous end-points and a 10% extra risk for quantal end-points.

Determining the uncertainty in each aspect is challenging, as was seen in section 4. Once that is done, however, the uncertainty calculations are relatively straightforward. As discussed in section 3, there are three options for uncertainty characterization: non-probabilistic, approximate probabilistic and full probabilistic, each of which is discussed below.

## A1.2 Non-probabilistic approach to combining uncertainties

In the non-probabilistic approach, only the LCL and the UCL are used, and the uncertainties are combined in a non-probabilistic fashion to estimate the extreme range of possible results. Therefore, from equation A1-1, the non-probabilistic lower and upper bounds would be:

$$LCL_{non-prob}(HD_M^I) = \frac{LCL(BMD_M)}{UCL(Aspect_1) \cdot UCL(Aspect_2) \cdot \dots \cdot UCL(Aspect_n)} \quad (A1-6)$$

$$UCL_{non-prob}(HD_M^I) = \frac{UCL(BMD_M)}{LCL(Aspect_1) \cdot LCL(Aspect_2) \cdot \dots \cdot LCL(Aspect_n)} \quad (A1-7)$$

Based on the results from section 4, the (LCL, UCL) for each uncertain quantity can be taken as its respective 90% confidence interval. If a BMD has been estimated from the data, the (LCL, UCL) would simply be the (BMDL, BMDU) interval. If only a NOAEL is available, no case-specific distribution can be derived. However, in this case, the LCL and UCL for the BMD can be estimated based on the generic distribution describing the uncertainty in using the NOAEL as a surrogate for the BMD (see section 4.2.2 and Table 4.6).

The result of this approach is a simple-to-calculate (LCL, UCL) interval for the target human dose  $HD_M^I$ . In terms of the two components of uncertainty introduced in section 2.4, the (LCL, UCL) interval can be thought of as an estimate of the “degree of uncertainty”. However, from a coverage point of view, this interval can be characterized only as being a “greater than 90% confidence interval”. The actual coverage will differ on a case-by-case basis, which is a limitation that is shared by the traditional approach using single-valued factors (and PoDs). The main advantage of the non-probabilistic approach compared with the

deterministic approach is that both the lower and upper confidence bounds are considered. The result is not very precise in terms of either coverage or degree of uncertainty, although even such imprecise uncertainty characterization may be adequate in some cases for use in risk assessment (e.g. if exposure is much lower than the LCL). However, a more precise characterization of uncertainty requires a probabilistic calculation.

### A1.3 Approximate probabilistic approach to combining uncertainties

Under the approximate probabilistic approach, each variable in equation A1-1 is assumed to be approximately lognormally distributed. Because all the components in equation A1-1 are multiplied together, use of lognormal distributions greatly simplifies the calculations. Moreover, the resulting target human dose,  $HD_M^I$ , is also a lognormal distribution. In section 4, each distribution was characterized by its median P50 value and P95/P50 ratio, or equivalently by its P05 and P95 values. If each variable in equation A1-1 is assumed to be approximately lognormally distributed, then the resulting approximate P50 and P95/P50 values for  $HD_M^I$  are:

$$P50_{\text{approx-prob}}(HD_M^I) = \frac{P50_{\text{approx}}(BMD_M)}{P50_{\text{approx}}(\text{Aspect}_1) \cdot P50_{\text{approx}}(\text{Aspect}_2) \cdot \dots \cdot P50_{\text{approx}}(\text{Aspect}_n)} \quad (A1-8)$$

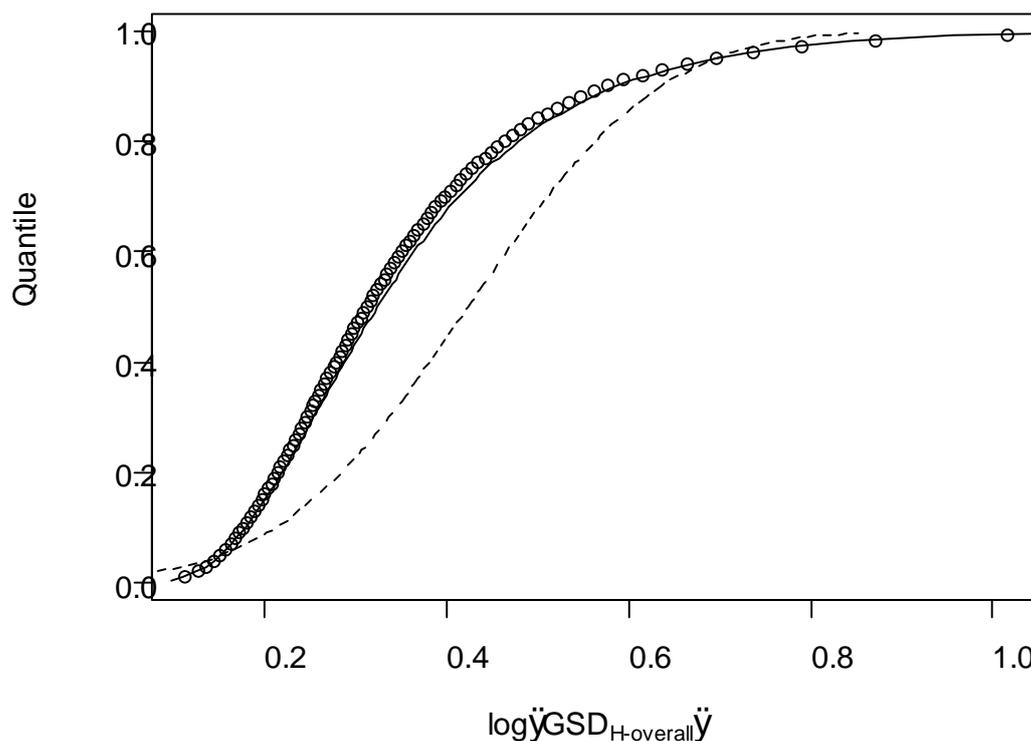
$$\frac{\ln \left( \frac{P95}{P50} \right)_{\text{approx-prob}}(HD_M^I)}{\ln \left( \frac{P95}{P50} \right)_{\text{approx}}(BMD_M)} = \frac{\ln \left( \frac{P95}{P50} \right)_{\text{approx}}(BMD_M)}{\ln \left( \frac{P95}{P50} \right)_{\text{approx}}(BMD_M)} + \sum_i \frac{\ln \left( \frac{P95}{P50} \right)_{\text{approx}}(\text{Aspect}_i)}{\ln \left( \frac{P95}{P50} \right)_{\text{approx}}(BMD_M)} \quad (A1-9)$$

The results of this approach give a more precise characterization of both the coverage and degree of combined uncertainty. For instance, one is 95% confident (coverage) that the true value of  $HD_M^I$  is greater than (less than) the LCL (UCL). Similarly, the interval (degree of uncertainty) with 90% confidence (coverage) can be estimated as having an LCL =  $P50/(P95/P50)$  and a UCL =  $P50 \cdot (P95/P50)$ .

For most of the aspects discussed in section 4, the uncertainty distribution is already assumed to be lognormal, so the P50 and P95/P50 values from Table 4.6 can be used to specify the distribution used in the calculations.

However, for two of the aspects discussed in section 4, the BMD and intraspecies variability, the underlying distribution is not necessarily (close to) lognormal. In such a case, there may be multiple ways of specifying the approximate lognormal distribution. In particular, one may approximate the true distribution with a lognormal distribution that has the same P50 and P95 values or one that has the same P05 and P95 values. These will give different answers, when the underlying distributions are not symmetric after logarithmic transformation (i.e. P95/P50 is not equal to P50/P05). For the approximate probabilistic approach, it was determined that matching P05 and P95 values would be preferred, as the main goal is to determine the overall confidence interval. Therefore, the BMD is approximated by a lognormal distribution with  $P05 = BMDL$  and  $P95 = BMDU$ . Similarly, the intraspecies variation (at a given target population incidence  $I$ ) is approximated by a lognormal distribution, with P05 calculated using the P05 of  $GSD_H$  and P95 calculated using the P95 of  $GSD_H$ .

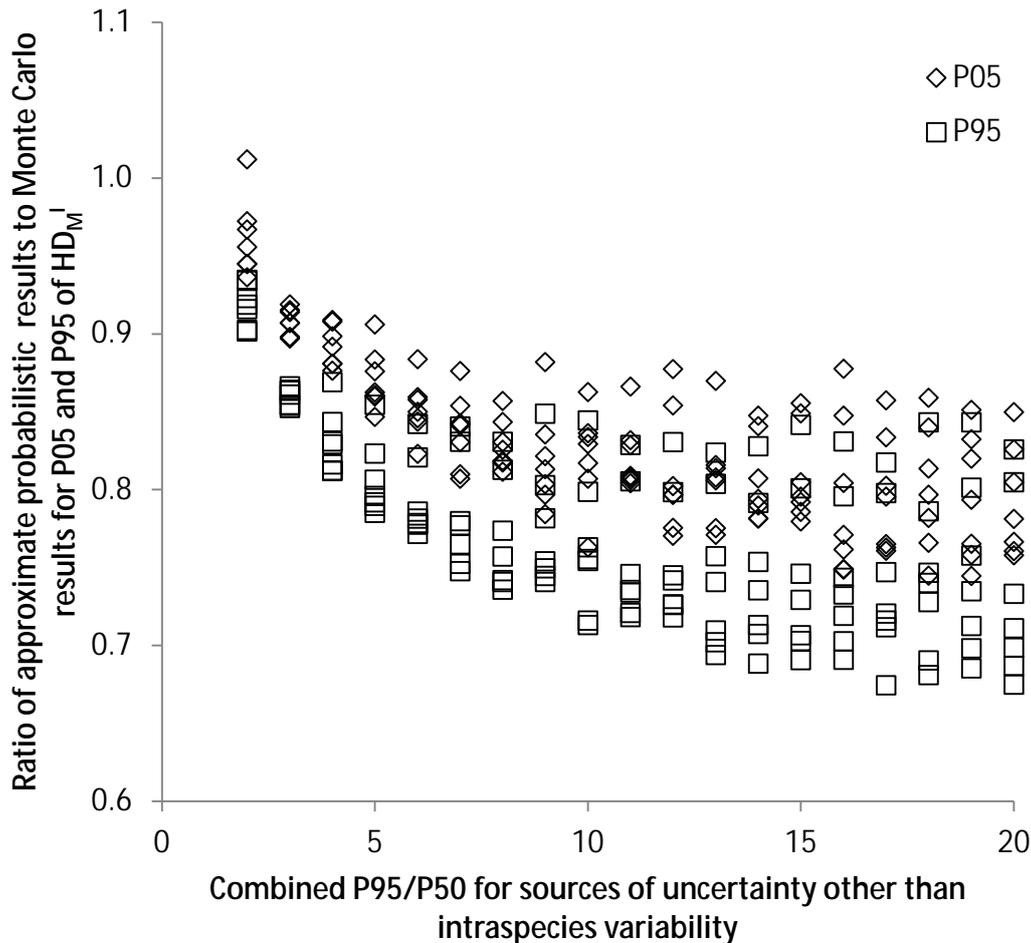
In contrast to the full probabilistic approach discussed below, the approximate probabilistic approach essentially assumes that  $\log(\text{GSD}_H)$  is normally distributed, rather than lognormally distributed, therefore introducing some error in the shape of the distribution. Fig. A1.1 shows the distribution of  $\log(\text{GSD}_H)$  generated from Monte Carlo simulation (see Annex 4) compared with both the lognormal approximation and the normal approximation, the latter of which is used by APROBA. Except in the extreme tails, the normal approximation is somewhat more conservative than the simulated distribution, in the sense that slightly larger values of  $\log(\text{GSD}_H)$  will be used.



**Fig. A1.1: Comparison of different approximations for the uncertainty distribution of  $\log(\text{GSD}_H)$ .** The circles were generated by Monte Carlo simulation (see Annex 4); the solid line represents a lognormal distribution, and the dashed line a normal distribution, both fitted by matching the same P05 and P95 values as in the Monte Carlo simulation.

Although calculations based on this approach are approximate, in simulations for the generic scenarios described in section 5, the differences in the P05 or P95 values between the approximate probabilistic approach and the Monte Carlo simulation were no more than 30% (see Table 5.1). Similar results were obtained over a larger range of possible values for the various uncertainties, as shown in Fig. A1.2. Specifically, in the scenarios run, the ratio of the results of the approximate probabilistic approach and Monte Carlo simulation for the P05 and P95 of the  $\text{HD}_M^I$  was between 0.67 and 1.01.

The approach of lognormal distributions for all components in equation A1-1 can be easily implemented in a spreadsheet to facilitate harmonized implementation of the approach in routine assessments. A prototype software tool has therefore been developed in conjunction with this document, called “APROBA”, for “Approximate PROBABILISTIC Analysis”, and is implemented in Microsoft Excel to ensure wide accessibility. Instructions on the use of APROBA are included in Annex 2.



**Fig. A1.2: Comparison of confidence limits on  $HD_M^I$  from approximate probabilistic approach and Monte Carlo simulation.** Specifically, the y-axis shows the ratio between the approximate probabilistic approach and Monte Carlo simulation for the P05 (diamonds) and P95 (squares) of the  $HD_M^I$ . Simulations included a range of values for the contribution from sources of uncertainty other than intraspecies variability (x-axis). The scatter in values along the y-axis corresponds to simulations using a range of the values for intraspecies variability, derived from values of incidence  $I$  ranging from 0.01% to 10%, using the preliminary distribution for intraspecies variability from section 4.

It is expected that, initially, the approximate probabilistic approach will be the most widely implemented probabilistic approach to characterizing uncertainty, as it provides a more precise estimate of uncertainty without requiring specialized software. The use of the APROBA tool is illustrated comprehensively in a specific case-study with deoxynivalenol, provided in Annex 5.

#### A1.4 Full probabilistic approach to combining uncertainties

Based on the information in Annexes 3 and 4, there is little reason to assume a distribution other than lognormal for the uncertainty in any of the components of equation A1-1 except for uncertainties related to (1) the BMD and (2) human intraspecies variability.

In terms of the BMD, the precise distributional shape will depend on the dose–response data set, as, in general, non-linear dose–response models are used to estimate the BMD. For human intraspecies variability, it is the  $\log(GSD_H)$  that is assumed to have a lognormal

uncertainty distribution (see section 4), whereas the approximate approach assumes that the interspecies factor itself, which is given by  $GSD_H^{z_i-1}$ , is lognormal.

In general, multiple distributions with different shapes cannot be combined in a simple mathematical formula that could be implemented in a spreadsheet. Therefore, a full probabilistic approach for calculating the uncertainty in the target human dose,  $HD_M^I$ , requires using Monte Carlo simulation. Specifically, random samples are generated for each component of equation A1-1, which are combined to form a set of random samples for  $HD_M^I$ . Then, the confidence interval on  $HD_M^I$  can be calculated from the sample quantiles of the random samples, thus providing an estimate of the degree of uncertainty for any desired per cent coverage.

This approach is more time-consuming, particularly because specialized software is needed to generate random samples from the BMD distribution. However, as mentioned above, the results of the full probabilistic approach are quite similar to the results from the approximate probabilistic approach in a number of simulations.

## ANNEX 2: APROBA SPREADSHEET TOOL USER GUIDE

### A2.1 Introduction

The purpose of the APROBA tool<sup>18</sup> is to facilitate non-probabilistic and approximate probabilistic uncertainty analysis, as described in section 3. The computational approaches are described in Annex 1 (sections A1.2 and A1.3) and not repeated here.

The APROBA tool is a Microsoft Excel workbook with three worksheets:

1. The worksheet “Wksht.LCL,UCL” performs non-probabilistic and approximate probabilistic analyses. The uncertainty in each hazard characterization aspect (PoD, Interspecies, etc.) is specified in terms of the 5% lower confidence limit (LCL = P05) and the 95% upper confidence limit (UCL = P95).
2. The worksheet “Provisional Parameter Values” contains the “standard” values for many of the inputs and uncertainties. These are either based on nominal default values (such as 60 kg for human body weight) or based on the generic uncertainties described and estimated in section 4, with details in Annexes 3 and 4.
3. The worksheet “Pick Lists” contains the allowed choices for some of the input variables.

All the worksheets are “locked”, so that most cells and formulas cannot be changed. The only cells that may be changed are those for which the user may enter inputs (highlighted in light yellow).

### A2.2 General layout of worksheets

Fig. A2.1 shows the general layout of the worksheet “Wksht.LCL,UCL”, showing the different parts of the worksheet. As shown in this figure, there are basically four parts to the worksheet:

1. inputs related to the study, end-point and protection goals;
2. inputs related to adjustment, variability and uncertainty (i.e. the aspects of hazard characterization);
3. intermediate calculations for uncertainty analysis; and
4. outputs.

### A2.3 Step-by-step procedure for using APROBA

The steps to using the APROBA tool are as follows:

---

<sup>18</sup> Note: This description refers to APROBA version 1.00. In the future, the tool might be developed further, which may result in slight discrepancies between the text and/or figures in this section and the most recent APROBA version.

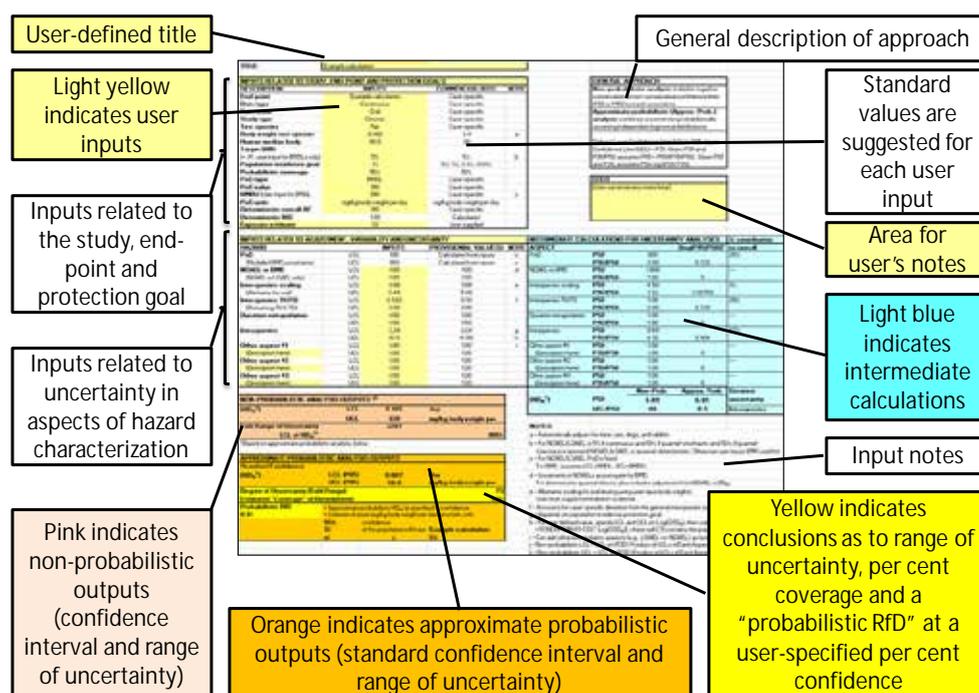


Fig. A2.1: General layout of main calculation section of APROBA worksheet “Wksht.LCL,UCL”

1. **Enter inputs related to the study, end-point and protection goals.** Detailed instructions are given in Table A2.1. If available, standard values are suggested next to each relevant input. These may be used, or the user may enter values specific to his or her situation. For instance, the user may desire a probabilistic coverage goal of greater than the standard value of 95%.
2. **Enter inputs related to adjustment, variability and uncertainty (i.e. the aspects of hazard characterization).** Detailed instructions are given in Table A2.2. If available, standard values are suggested next to each relevant input. These may be used, or the user may enter values specific to his or her situation. For instance, if there are chemical-specific data, such as a PBTK model, the standard interspecies scaling based on allometric scaling may be altered to the value appropriate for the particular compound and end-point being characterized.
3. **Extract outputs and conclusions.** The spreadsheet will automatically calculate the following outputs and quantitative conclusions:
  - a. **Non-probabilistic LCL and UCL, and associated fold range of uncertainty.** The only statement that can be made from those outputs alone is that the (LCL, UCL) interval has “more than 95% coverage”.
  - b. **“Coverage” of the non-probabilistic LCL.** This output is the per cent confidence, based on the approximate probabilistic analysis, that the actual target human dose,  $HD_M^1$ , is greater than the LCL derived from the non-probabilistic analysis.
  - c. **Approximate probabilistic LCL and UCL, and associated fold range of uncertainty.** These outputs give an approximate 95% confidence interval and the associated degree of uncertainty.
  - d. **“Coverage” of the original RfD.** This output is the per cent confidence, based on the approximate probabilistic analysis, that the actual target human dose,  $HD_M^1$ , is greater than the original RfD.

**Table A2.1: Detailed instructions for inputs related to study, end-point and protection goals**

<i>Row description</i>	<i>Instructions</i>
End-point	Enter a description of the end-point.
Data type	Choose from a drop-down list with choices: <ul style="list-style-type: none"> <li>• “Continuous” for continuous end-points.</li> <li>• “Quantal-deterministic” for “deterministic” quantal end-points, where the observed dose–response relationship represents experimental variation (e.g. histological end-points).</li> <li>• “Quantal-stochastic” for “stochastic” quantal end-points, where the observed dose–response relationship represents an individual probability of developing the end-point, such as cancer or malformations.</li> </ul> <p>For reference, see section 3.</p>
Data route	Choose from a drop-down list that includes “Oral”, “Inhalation” and “Dermal”. The modules for other exposure routes have not yet been developed.
Study type	Choose from a drop-down list with choices “Chronic”, “Subchronic”, “Subacute” or “Repro/Developmental”.
Test species	Choose from a drop-down list with choices “Rat” or “Mouse”.
Body weight test species (kg)	Enter manually based on that reported in the study. “Standard values” are provided for reference.
Human median body weight (kg)	Enter manually based on the human population whose risk is being assessed. A “standard value” of 60 kg is provided for reference.
Target BMR (= <i>M</i> , user input for BMDLs only)	If BMD modelling is performed, then enter the BMR here. “Standard values” of 5% relative change for continuous end-points, 50% extra risk for deterministic quantal end-points and 10% extra risk for stochastic quantal end-points are provided for reference. If a NOAEL is being used as the PoD, then the “standard value” is appropriate.
Population incidence goal (= <i>I</i> )	Enter the target population incidence – i.e. the fraction of the population for whom an effect of magnitude equal to the “Target BMR” would be acceptable. “Standard values” may be 5%, 1%, 0.1% or 0.01%.
Probabilistic coverage goal	Enter the per cent confidence (“coverage”) desired in the final probabilistic result. A “standard value” is 95%.
PoD type	Choose from a drop-down list with choices “NOAEL” or “BMD”.
PoD value	Enter the numerical value of the PoD used in the original RfD calculation (e.g. the NOAEL value or the BMDL value).
BMDU (user input for BMDL PoDs)	If the PoD is a BMDL, then enter the numerical value of the BMDU derived from BMD modelling. Leave blank if PoD is NOAEL.
PoD units	Enter the units of the PoD, such as “mg/kg body weight per day”.
Deterministic overall AF	Enter the overall (or “composite”) assessment factor (or “uncertainty factor”) used to calculate the Deterministic RfD (in the next row) = PoD value / Deterministic overall AF.
Deterministic RfD	Not user input – calculated as PoD value / Deterministic overall AF.
Exposure estimate	An optional input for comparing an exposure value with an RfD (the description “Exposure estimate” can also be changed). Also, this can appear in the graphical display (see section A2.5).

- e. **Probabilistic RfD.** This output is the LCL of the target human dose,  $HD_M^I$ , at the user-specified per cent confidence, based on the approximate probabilistic analysis. It is described as the “Estimate of the dose [units] at which, with [user-specified per

cent] confidence, [user-specified population incidence goal per cent] of the population will have [user-entered end-point] of magnitude greater than [user-specified target BMR].

**Table A2.2: Detailed instructions for inputs related to adjustment, variability and uncertainty**

<i>Row description</i>	<i>Instructions</i>	
PoD	LCL UCL	This aspect addresses uncertainty in the PoD. These are automatically calculated based on the previous user inputs. <ul style="list-style-type: none"> <li>For a PoD that is a BMDL, LCL = BMDL and UCL = BMDU.</li> <li>For a PoD that is a NOAEL, the PoD is fixed and LCL = UCL = NOAEL.</li> </ul>
NOAEL to BMD	LCL UCL	This aspect addresses the uncertainty of using a NOAEL as an estimate of the BMD. <ul style="list-style-type: none"> <li>For a PoD that is a BMDL, this aspect is not included, and both values should be set equal to 1.</li> <li>For a PoD that is a NOAEL, standard values based on historical data as described in section 4 (listed in Table 4.6) are suggested, but the user can enter a different value.</li> </ul>
Interspecies scaling	LCL UCL	This aspect addresses the interspecies adjustment to take into account differences in body size. Standard values for allometric scaling as described in section 4 (listed in Table 4.6) are suggested, but the user can enter a different value.
Interspecies TK/TD	LCL UCL	This aspect addresses remaining interspecies TK and TD differences after accounting for body size differences. Standard values based on historical data as described in section 4 (listed in Table 4.6) are suggested, but the user can enter a different value.
Duration extrapolation	LCL UCL	This aspect addresses uncertainty in using a less-than-chronic study (as specified in “Study type” previously) to estimate a chronic PoD. Standard values based on historical data as described in section 4 (listed in Table 4.6) are suggested, but the user can enter a different value.
Intraspecies	LCL UCL	This aspect addresses the uncertainty in the amount of human variability in sensitivity. It depends directly on the “population incidence goal” entered previously (described in Table A2.1). Standard values based on historical data as described in section 4 (listed in Table 4.6) are suggested, but the user can enter a different value. Note that if the user has a different suggested LCL and UCL of the intraspecies variability $\log(\text{GSD}_H)$ , he/she will need to calculate the LCL and UCL of $\text{AF}_{\text{intra-}I}$ associated with the specified incidence $I$ using the formula $\text{AF}_{\text{intra-}I} = 10^{z_{1-I} \log(\text{GSD}_H)}$ , where $z_{1-I}$ is the z-score for normal distribution corresponding to a percentile $1 - I$ .
Other aspect #1/#2/#3	LCL UCL	If there are other aspects of hazard characterization that need to be incorporated, they can be added by the user in these rows.

## A2.4 Using intermediate calculations to estimate each aspect's contributions to uncertainty

The “intermediate calculations” shown in the APROBA worksheets may give useful results in terms of deciding whether to conduct additional analysis, modelling or data generation. In particular, the column marked “[log(P95/P50)]<sup>2</sup>” gives the contribution from each aspect to the overall log variance. Thus, the highest values in this column correspond to the aspect of hazard characterization that contributes most to the uncertainty. Additionally, a column marked “% contribution to overall uncertainty” gives the percentage of the overall [log(P95/P50)]<sup>2</sup> in HD<sub>M</sub><sup>I</sup> that is contributed by each hazard characterization aspect. This information can help prioritize efforts to reduce uncertainty. For instance, if NOAEL to BMDL uncertainty is among the greatest sources of uncertainty, then conducting BMD modelling could significantly reduce uncertainty. If the data set is not amenable to such modelling, then additional experiments designed so that BMD modelling is feasible could be a priority. Alternatively, if duration extrapolation is among the greatest sources of uncertainty, then additional longer-duration studies may significantly reduce uncertainty. Overall, however, the “intermediate calculations” provide valuable insight from the probabilistic calculations into the relative contributions of different sources of uncertainty.

## A2.5 Visualizing the impact of changing coverage and incidence *I*

The APROBA tool also includes the capability of graphically displaying the impact of different choices for coverage and incidence *I* on the estimate of the HD<sub>M</sub><sup>I</sup>, as well as a comparison with the deterministic and probabilistic RfDs calculated previously and with the exposure estimate previously entered (see Table A2.1). The user inputs that allow for the graphical display of the impact of changing coverage and incidence *I* are described in Table A2.3, with a screenshot of the area of the workbook where this capability is controlled shown in Fig. A2.2. The format of the graphic (axes, colours, line styles, etc.) itself can be modified per standard features in Microsoft Excel.

**Table A2.3: Detailed instructions for inputs related to graphical display**

<b><i>Row description</i></b>	<b><i>Instructions</i></b>
Min incidence ( <i>I</i> ) shown	The user enters the lowest and highest values of population incidence <i>I</i> that are to be displayed in the graphic. The standard values suggested are 0.01% and 50%.
Max incidence ( <i>I</i> ) shown	
Coverage percentages shown (decending order)	The user enters different levels of coverage that are to be displayed in the graphic. The standard values suggested are 99%, 95%, 90%, 10%, 5% and 1%.

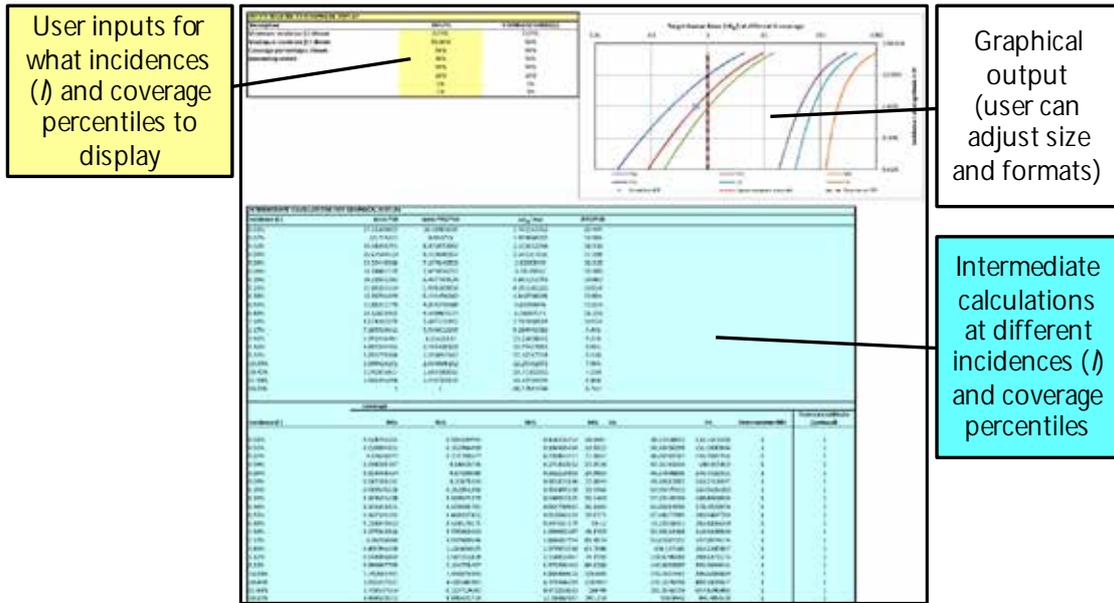


Fig. A2.2: General layout of graphical display section of APROBA worksheet "Wksht.LCL,UCL"

## ANNEX 3: DETAILS OF ESTIMATING UNCERTAINTY DISTRIBUTIONS FOR USE OF A NO-OBSERVED-ADVERSE-EFFECT LEVEL, EXPOSURE DURATION, INTERSPECIES EXTRAPOLATION AND ROUTE-TO-ROUTE EXTRAPOLATION

### A3.1 Point of departure: uncertainty in using a NOAEL as a surrogate for a BMD

#### **A3.1.1 Continuous end-points**

EFSA (2009) reports a NOAEL/BMDL ratio distribution related to 395 dose–response data sets from (oral subchronic and chronic toxicity) studies conducted by the United States National Toxicology Program (NTP). These relate to various end-points (body weight, liver weight, kidney weight, red blood cell count), but no differences in NOAEL/BMDL distributions among these end-points were apparent. The (geometric) mean ratio was around 1 for a BMR of 5%. The P95 differed from the GM by a factor of about 3.

Allen et al. (1994) reviewed a large number of (oral) developmental toxicity studies, focusing on the fraction of affected fetuses or implants per litter as a continuous<sup>19</sup> end-point. They reported that the median ratio of NOAEL to BMDL (note that the authors used the term BMD for BMDL) was around 1, whereas 95% of the NOAELs were within a factor of 5 of the associated BMDL (at BMR = 5%).

Based on the previous discussion, the following default distributions might be considered as reasonably reflecting the uncertainty in a NOAEL as an estimate of the BMDL<sub>05</sub>, depending on study type:

(Oral) chronic/subchronic toxicity studies:      GM = 1, with P95/P50 = 3

(Oral) developmental toxicity studies:      GM = 1, with P95/P50 = 5

As no data related to other exposure routes were found, an assumption of similar distributions for other routes may be considered.

As these distributions are for a NOAEL as an estimate of the BMDL, additional uncertainty in the BMD itself needs to be accounted for in order to estimate the overall uncertainty in the true BMD. Obviously, this uncertainty (reflected by the BMDU/BMDL ratio) depends on the specific data. In principle, a systematic reanalysis of historical data could provide a distribution of BMDU/BMDL ratios, which could then be used as an uncertainty distribution. This would probably show that the uncertainty in the BMD is often larger for quantal data than for continuous data (as quantal data are less informative than continuous data). For the time being, based on general experience, it will be assumed that the BMDU/BMDL ratio is about 9, as a mildly conservative value. Thus, the GM of 1 of the uncertainty distributions just mentioned would be divided by 3 to arrive at the true BMD, rather than the BMDL. Further, the factor P95/P50 of these distributions needs to be inflated by an additional

---

<sup>19</sup> Strictly speaking, this is not a continuous end-point, but was regarded as such by the authors of that study.

P95/P50 = 3, reflecting the assumed uncertainty in the BMD (see Annex 1 for the formula for combining uncertainties).

The resulting overall uncertainty distributions for a NOAEL for a continuous end-point are as follows:

(Oral) chronic/subchronic toxicity studies:

$$GM = 1/3, \text{ with } P95/P50 = 4.7 \text{ [(P05, P95) = (0.07, 1.6)]}$$

(Oral) developmental toxicity studies:

$$GM = 1/3, \text{ with } P95/P50 = 7.0 \text{ [(P05, P95) = (0.05, 2.3)]}$$

### **A3.1.2 Quantal end-points**

In their reanalysis of developmental toxicity studies, Allen et al. (1994) also reported NOAEL/ BMDL ratios for a quantal end-point (frequency of litters affected). For BMR = 10%, the median NOAEL/BMDL ratio was about 2, whereas 88% of the ratios were lower than 5, and the P95 would have been somewhat larger, perhaps 6.

Based on the previous discussion, the following default distribution might be considered as reasonably reflecting the uncertainty in a NOAEL as an estimate of the BMDL<sub>10</sub>:

(Oral) developmental toxicity studies: GM = 2, with P95/P50 = 3

Just as in the case of continuous end-points, an additional uncertainty distribution is needed to translate uncertainty in the BMDL<sub>10</sub> into uncertainty in the BMD. The latter uncertainty distribution may again be assumed to have GM = 1/3 and P95/P50 = 3.

In the case of quantal end-points that are considered to be “deterministic” (e.g. histopathological end-points), there is one additional complication. As discussed in section 3.2, in these end-points, the target human dose, HD<sub>M</sub><sup>I</sup>, would be defined with *M* being the severity category of the lesion, and the appropriate PoD would be the ED<sub>50</sub> of the dose–response data in the test animal. However, the NOAEL is regarded as an estimate of the BMDL<sub>10</sub>. Therefore, an additional assessment factor needs to be applied, reflecting the assumed distance between the BMD<sub>10</sub> and the ED<sub>50</sub>. From a reanalysis of historical data, it is known that the distance between the BMD<sub>10</sub> and the ED<sub>50</sub> is fairly constant in quantal dose–response relationships (W. Slob & R.W. Setzer, unpublished data, 2014), with mild variation around the value of 3. Therefore, for these quantal end-points, an additional uncertainty distribution needs to be applied, with GM = 1/3 and P95/P50 = 1.5.

This results in the following overall uncertainty distribution for a quantal NOAEL:

Developmental toxicity studies, deterministic quantal end-points:

$$GM = 2/9, P95/P50 = 5 \text{ [(P05, P95) = (0.04, 1.1)]}$$

Developmental toxicity studies, stochastic quantal end-points:

$$GM = 2/3, P95/P50 = 4.7 \text{ [(P05, P95) = (0.14, 3.2)]}$$

For other study types, no useful data were found. With this lack of information, use of a distribution with the same GM but a larger P95/P50 might be considered, to reflect the additional uncertainty.

## A3.2 Exposure duration: uncertainty in using a subchronic or subacute study as a surrogate for a chronic study

### A3.2.1 Subchronic to chronic extrapolation

For the uncertainty distribution related to the subchronic to chronic extrapolation, several studies comparing oral NOAELs from chronic and subchronic toxicity studies appear relevant and are summarized in Table A3.1 for oral exposures and in Table A3.2 for inhalation exposures.

**Table A3.1: Subchronic to chronic oral NO(A)EL ratios**

<i>N</i>	<i>GM</i>	<i>P95/P50<sup>a</sup></i>	<i>Subchronic exposure period</i>	<i>Chronic exposure period</i>	<i>Species</i>	<i>Reference</i>
33	2.2	4	30–210 days	2 years	Rats	Weil & McCollister (1963)
41	1.0	2.5	Not specified	Not specified	Rats, dogs <sup>b</sup>	McNamara (1976)
20	1.9	6	< 200 days	> 200 days	Various	Rulis & Hattan (1985)
149	1.7	17	10–26 weeks	1–2 years	Various rodents <sup>c</sup>	Pieters, Kramer & Slob (1998)
23	2.0	2.5	90 days	2 years	Various rodents <sup>c</sup>	Nessel et al. (1995)
9	2.4	1.5	90 days	1–2 years	Mice	Kalberlah & Schneider (1998) <sup>d</sup>
11	1.7	2.5	90 days	1–2 years	Rats	Kalberlah & Schneider (1998) <sup>d</sup>
20	2.0	4	90 days	1–2 years	Mice + rats	Kalberlah & Schneider (1998) <sup>d</sup>
21	1.7	2	90 days	2 years	Mice	Kalberlah & Schneider (1998) <sup>e</sup>
22	2.5	3	90 days	2 years	Rats	Kalberlah & Schneider (1998) <sup>e</sup>
68	1.5	15	90 days	2 years	Mice + rats	Bokkers & Slob (2005)
70	2.25	8	49–183 days	≥ 1 year	Mice + rats	Groeneveld et al. (2004)
56	2.28	9	49–183 days	≥ 1 year	Rats	Groeneveld et al. (2004)
236	1.5	9.8	83–99 days	> 699 days	Mice + rats	Batke et al. (2011) <sup>f</sup>
58	1.4	3.3	83–99 days	> 699 days	Mice + rats	Batke et al. (2011) <sup>g</sup>

GM: geometric mean; GSD: geometric standard deviation; LOAEL: lowest-observed-adverse-effect level; *N*: number of ratios; NO(A)EL: no-observed-(adverse-)effect level; NTP: (United States) National Toxicology Program; P50: 50th percentile; P95: 95th percentile

<sup>a</sup> Percentiles calculated from the GM and GSD.

<sup>b</sup> Thirty-nine rat pairs, two dog pairs.

<sup>c</sup> Matched pairs.

<sup>d</sup> Industry data from 13 agrochemicals.

<sup>e</sup> Data from the NTP.

<sup>f</sup> Pairs matched for species, NO(A)ELs + LO(A)ELs, total database.

<sup>g</sup> Studies of comparable design, pairs matched for species, NO(A)ELs only.

Table A3.2: Subchronic to chronic inhalation NOAEL ratios

<i>N</i>	<i>GM</i>	<i>P90/P50</i>	<i>Species</i>	<i>Reference</i>
12	2.8	3.9	Rats	Kalberlah, Föst & Schneider (2002) <sup>a</sup>
16	3.3	6.7	Mice	Kalberlah, Föst & Schneider (2002) <sup>a</sup>
68	2.7	7.4	Mice + rats	Kalberlah, Föst & Schneider (2002) <sup>a,b</sup>
101	1.6	5.6	Mice + rats	Batke et al. (2011) <sup>c,d</sup>
19	2.1	2.8	Mice + rats	Batke et al. (2011) <sup>c,e</sup>

GM: geometric mean; LOAEL: lowest-observed-adverse-effect level; *N*: number of ratios; NOAEL: no-observed-adverse-effect level; NTP: (United States) National Toxicology Program; P50: 50th percentile; P90: 90th percentile (95th percentiles were not reported by the authors)

<sup>a</sup> NTP studies, subchronic = 90 days, chronic = 2 years.

<sup>b</sup> Includes both NOAEL and LOAEL ratios.

<sup>c</sup> Subchronic = 83–99 days, chronic = > 699 days.

<sup>d</sup> Pairs matched for species, NOAELs + LOAELs, total database.

<sup>e</sup> Studies of comparable design, pairs matched for species, NOAELs only.

It is very likely that the databases used in the various studies overlap each other substantially. Some of them used secondary sources, but those that used the primary data may have differed in the interpretation of the tests available.

In two of these studies (Rulis & Hattan, 1985; Pieters, Kramer & Slob, 1998), the ratios have not been matched for the species concerned, and one may expect broader distributions here because they partly include interspecies differences.

Bokkers & Slob (2005) reported oral subchronic to chronic BMD ratios, next to NOAEL ratios (for the same database). For the BMD ratios, they found a median of 1.7, with a P95/P50 ratio of about 4. Again, the variation in NOAEL ratios was found to be larger, the P95 value being about 15 times higher than the median (P50). However, various studies summarized in Table A3.1 reported factors lower than 4, the factor found for the BMD ratios in Bokkers & Slob (2005). This is remarkable, as higher values would be expected because (1) they relate to NOAEL ratios, (2) the latter studies do not match end-points, whereas Bokkers & Slob (2005) did, and (3) some of them did not even match species. In contrast, for example, Batke et al. (2011) calculated their ratios based on a millimole per kilogram body weight per day scale (oral studies), which can be expected to result in a narrower distribution than using milligrams per kilogram body weight per day, given that molecular masses of commonly used industrial chemicals span a range of 3–4 orders of magnitude.<sup>20</sup>

Baird et al. (1996) reported NOAEL ratios based on two pooled data sets with a total of 51 NOAEL ratios of both oral and inhalation studies, with GM = 2.1 and P95/P50 = 4.5.

The reported GM of the oral subchronic to chronic ratios varies between 1 and 2.5, whereas most of them are close to 2. The width of the distribution varies considerably among the studies, which is not clearly understood, although variability in study design is an important factor. When taking the reported BMD ratios (Bokkers & Slob, 2005) as the most relevant information, the P95/P50 ratio would be 4. For inhalation, no clear evidence was found that the distribution is really different. So, a choice for the uncertainty distribution of the subchronic to chronic factor for oral or inhalation doses might be:

<sup>20</sup> For this report, molecular mass data from the European Inventory of Existing Chemical Substances were analysed with descriptive statistics (using SigmaPlot 11 software); *N* = 72 356, range = 1–3500 g/mol, (arithmetic) mean = 315 g/mol, median = 252 g/mol, SD = 205 g/mol, P25 = 18, P75 = 386, as published under [http://ihcp.jrc.ec.europa.eu/our\\_labs/predictive\\_toxicology/information-sources/ec\\_inventory/einecs\\_100204\\_19July06.zip](http://ihcp.jrc.ec.europa.eu/our_labs/predictive_toxicology/information-sources/ec_inventory/einecs_100204_19July06.zip).

Subchronic to chronic extrapolation:

$$GM = 2, \text{ with } P95/P50 = 4 \text{ [(P05, P95) = (0.5, 8)]}$$

**A3.2.2 Subacute to chronic extrapolation**

Tables A3.3 and A3.4 summarize a number of studies that reported subacute to chronic NOAEL ratios. No BMD ratio studies are available. The study by Kramer et al. (1996) did not match species.

**Table A3.3: Subacute to chronic oral NO(A)EL ratios**

<i>N</i>	<i>GM</i>	<i>P95/P50<sup>a</sup></i>	<i>Subacute exposure period</i>	<i>Chronic exposure period</i>	<i>Species</i>	<i>Reference</i>
71	4.1	11	3–6 weeks	1–2 years	Various	Kramer et al. (1996)
20	3.1	3	14 days	2 years	Mice	Kalberlah & Schneider (1998) <sup>b</sup>
26	3.9	4	14 days	2 years	Rats	Kalberlah & Schneider (1998) <sup>c</sup>
35	4.88	8	21–42 days	≥ 1 year	Mice + rats	Groeneveld et al. (2004)
25	5.81	8	21–42 days	≥ 1 year	Rats	Groeneveld et al. (2004)
49	2.9	13.6	20–33 days	> 699 days	Mice + rats	Batke et al. (2011) <sup>d</sup>
14	3.4	8.6	20–33 days	> 699 days	Mice + rats	Batke et al. (2011) <sup>e</sup>

GM: geometric mean; GSD: geometric standard deviation; LOEL: lowest-observed-effect level; *N*: number of ratios; NO(A)EL: no-observed-(adverse-)effect level; NTP: (United States) National Toxicology Program; P50: 50th percentile; P95: 95th percentile

<sup>a</sup> Percentiles calculated from the GM and GSD.

<sup>b</sup> Industry data from 13 agrochemicals.

<sup>c</sup> Data from the NTP.

<sup>d</sup> Pairs matched for species, NOELs + LOELs, total database.

<sup>e</sup> Studies of comparable design, pairs matched for species, NO(A)ELs only.

**Table A3.4: Subacute to chronic inhalation NOAEL ratios<sup>a,b</sup>**

<i>N</i>	<i>GM</i>	<i>P95/P50</i>	<i>Subacute exposure period</i>	<i>Chronic exposure period</i>	<i>Species</i>	<i>Reference</i>
13	3.2	3.7	14 days	2 years	Rats	Kalberlah, Föst & Schneider (2002)
10	7.0	4.9	14 days	2 years	Mice	Kalberlah, Föst & Schneider (2002)
59 <sup>c</sup>	7.2	2.9	14 days	2 years	Mice + rats	Kalberlah, Föst & Schneider (2002)
18	2.4	4.2	20–33 days	> 699 days	Mice + rats	Batke et al. (2011)

GM: geometric mean; GSD: geometric standard deviation; LOAEL: lowest-observed-adverse-effect level; *N*: number of ratios; NOAEL: no-observed-adverse-effect level; NTP: (United States) National Toxicology Program; P50: 50th percentile; P95: 95th percentile

<sup>a</sup> Percentiles calculated from the GM and GSD.

<sup>b</sup> Data from the NTP.

<sup>c</sup> Includes both NOAEL and LOAEL ratios.

The GMs of the oral subacute to chronic ratios are higher than those of the subchronic to chronic ratios and vary between 3 and 6. The factor P95/P50 varies between 3 and 14.

When omitting Kramer et al. (1996) for not matching species, the three remaining oral studies still deviate from each other. It is not clear which of these would be the most relevant.

For inhalation, the data are more limited. There is no clear indication of whether or not the oral and inhalation distributions are likely to be different from each other. Hence, a reasonable choice for the subacute to chronic distribution (oral and inhalation) might be:

Subacute to chronic extrapolation:

$$GM = 5, \text{ with } P95/P50 = 8 \text{ [(P05, P95) = (5/8, 40)]}$$

For studies with oral administration, this choice appears to be a conservative value, whereas for inhalation exposure, this is not known.

### A3.3 Interspecies extrapolation: uncertainty in using a test animal species as a surrogate for humans (after body size adjustment)

As discussed in section 4.4, interspecies extrapolation may be subdivided into two parts:

1. adjustment of the dose for (generic) differences in body size between test animal and humans; and
2. accounting for potential (chemical-specific) TK/TD differences.

Details regarding adjustment for generic body size differences were discussed in section 4.4.1 of the main text. This section of Annex 3 discusses the data on TK/TD differences, reviewing studies that compared PoD ratios among different pairs of animal species.

Vermeire et al. (1999) and Rennen et al. (2001) reported NOAEL ratios for various pairs of species (see Table A3.5). The NOAEL ratios for the pairs of species considered tend to approach the value of 1 when doses are allometrically scaled. This would imply that after allometric scaling, species are, on average over chemicals, equally sensitive. Bokkers & Slob (2007) reported similar findings for NOAEL ratios and BMD ratios derived from studies in rats and mice conducted by the NTP: after allometric scaling, the median ratio was close to 1. This is consistent with the approach taken here to separate the generic physiological differences related to body size from chemical-specific TK/TD differences.

The P95/P50 ratios in Table A3.5 are rather different between the two studies, in particular for the respiratory NOAELs, but they are fairly similar across pairs of species (in the same study). The data in Table A3.5 suffer from the fact that they are based on NOAELs and not BMDs and thus contain an extra element of uncertainty, resulting in imprecise estimates of the P95/P50 that are biased upwards. Further, no distinction was made between end-points, type of oral dosing and other factors that may cause additional scatter in the outcomes.

Bokkers & Slob (2007) compared rats and mice and found P95/P50 values of 13 and 19 for the NOAEL ratios in females and males, respectively. These values are in the range of those in Table A3.5, even though they used a more homogeneous data set (NTP studies only), whereas they matched the end-point in both species in each case.

Table A3.5: NOAEL ratios related to pairs of species

<i>Species pairs in NOAEL ratio</i>	<i>N</i>	<i>GM</i>	<i>P95</i>	<i>P95/P50</i>
<b>Oral route</b>				
Rat/dog	63	1.3	18.8	14
	71	2.3	27	12
Rat/dog (allometrically scaled)	63	0.5	6.6	13
	71	0.8	9	11
Mouse/rat	67	4.2	73.9	18
	78	3.2	37	12
Mouse/rat (allometrically scaled)	67	2.4	42.2	18
	78	1.9	21	11
Mouse/dog	40	6.4	124.6	19
	20	5.9	50	8
Mouse/dog (allometrically scaled)	40	1.3	24.9	19
	20	1.2	10	8
<b>Respiratory route</b>				
Mouse/rat	21	3.1	91.8	30
	19	1.5	11	7

GM: geometric mean; *N*: number of pairs; P50: 50th percentile; P95: 95th percentile

Source: Vermeire et al. (1999, upper number in each row); Rennen et al. (2001, lower number in each row)

Bokkers & Slob (2007) also reported distributions of BMD ratios for rats and mice for the end-points body weight, absolute/relative liver weight, absolute/relative kidney weight and red blood cell count. The P95/P50 values were not only lower (about 3) than those for the NOAEL ratios (based on the same data), they were also very similar among different end-points, as well as between species and sexes. The latter indicates that the differences in P95/P50 values as found for the NOAEL ratios are caused by the relatively weak information in NOAELs.

The factor of 3 that was found for the P95/P50 based on the BMD ratios might seem low relative to the current default interspecies factor of 10, but it should be noted that this factor of 3 holds for the allometrically scaled oral dose. When, for example, the mouse is the test species, the allometric scaling factor is about 10. The median of the interspecies distribution would already be close to 10, and the overall interspecies factor with 95% coverage would be equal to 30.

The studies just discussed do not include human data. A study by Price, Keenan & Swartout (2008) does include human data, but rather than ratios of PoDs, it reviewed ratios of acute and subacute maximum tolerated doses of cancer chemotherapeutic agents for humans compared with various animal species. A summary of results is shown in Table A3.6. Again, allometric scaling had the effect of bringing the median ratios back to approximately 1. The P90 differed from the median (P50) by a factor of between 3 and 5. Given the limitations of this study, it does not directly inform the uncertainty distribution of an interspecies factor in hazard characterization, but the results are roughly in line with those just discussed (and did not include human data), in particular regarding allometric scaling.

**Table A3.6: Summary statistics for species-specific distributions of ratios of maximum tolerated doses**

	<i>Mouse MTD / human MTD</i>	<i>Rat MTD / human MTD</i>	<i>Monkey MTD / human MTD</i>	<i>Dog MTD / human MTD</i>
Number of agents	54	17	34	56
Median (range)	7.7 (6.8–9.3)	3.0 (1.9–5.8)	2.5 (2.1–3.3)	1.0 (0.7–1.5)
Allometric scaling <sup>a</sup>	1.0 (1.1–1.26)	0.61 (0.39–1.2)	1.1 (0.95–1.5)	0.6 (0.41–0.89)
P95/P50	3.2	5.3	2.7	4.4

CI: confidence interval; MTD: maximum tolerated dose; P50: 50th percentile; P95: 95th percentile

<sup>a</sup> The authors allometrically scaled the values assuming an allometric power of 0.75 and values for species' body weights from Freireich et al. (1966).

Source: Price, Keenan & Swartout (2008)

For inhalation PoDs, the NOAEL ratios reported by Vermeire et al. (1999) and Rennen et al. (2001) are rather inconsistent between the two studies (see Table A3.5); in particular, the P95/P50 reported by Vermeire et al. (1999) is relatively high, whereas that reported by Rennen et al. (2001) is more consistent with the value from the oral NOAEL ratios. It would be worthwhile to investigate the origin of this large difference<sup>21</sup> and to expand the number of data sets analysed. Also, interspecies differences might depend on the nature of the material inhaled – for example, whether it is a soluble gas or a particle – and the type of effect (systemic or portal of entry); in retrospect, these considerations could have been addressed by applying RDDRs or RGDRs (see section 4.4.1) prior to calculating PoD ratios. Based on the limited database available, it is unclear whether the interspecies difference for inhalation PoDs is really different from that associated with oral NOAELs. Further research is urgently needed to generate more reliable information.

As discussed above, BMD ratios are typically better than NOAEL ratios for informing the interspecies uncertainty distribution. Therefore, the data from Bokkers & Slob (2007) might be preferred over the data provided in Table A3.5. A limitation is that these authors considered a limited number of end-points from repeated-dose studies, and these were all NTP studies. Thus, use of these data would assume that similar interspecies differences would occur for other end-points, such as developmental end-points, and for a much broader universe of chemicals than tested in these NTP studies. A distribution for TK/TD differences based on the Bokkers & Slob (2007) analysis might be:

TK/TD uncertainty after accounting for body size differences:

$$GM = 1, P95/P50 = 3 [(P05, P95) = (1/3, 3)]$$

#### A3.4 Route-to-route extrapolation: uncertainty in using an oral study as a basis for inhalation hazard, or vice versa

Pepelko & Withey (1985) reported ratios between the oral median lethal dose (LD<sub>50</sub>) and inhalation median lethal concentration (LC<sub>50</sub>) of 0.1–55 (assuming 100% absorption via both routes), based on available acute lethality data for 49 chemicals. Rennen et al. (2004) analysed several databases of multiple-dose studies and found pairs of oral and inhalation

<sup>21</sup> When the P95 was estimated by simply taking the P95 of the sample, the resulting value was highly sensitive to outliers (given the sample size of 21). Estimating the P95 by fitting a lognormal distribution would be much less sensitive to outliers.

studies for only 28 out of 215 substances searched. They reported the ratios of NOAELs (assuming 100% absorption via both routes) to be 0.03–326. Even fewer data were available for oral to dermal extrapolation. Overall, Rennen et al. (2004) concluded that substantially more experimental data would be necessary to derive an assessment factor for route-to-route extrapolation. Analyses so far do not seem to have matched data for end-points, probably due to lack of data.

Pepelko & Withey (1985) suggested that chemical-specific data are necessary for reliable route-to-route extrapolation. Further, they identified five major factors that may result in apparent differences in toxicity between routes:

- (1) differences in absorption efficiency; (2) differences in systemic effects; (3) occurrence of critical toxicological effects at the portal of entry; (4) first-pass effects resulting in inactivation or activation of the chemical agent before it reaches the target organ; and (5) variations in temporal patterns of target organ concentrations.

The analyses of Rennen et al. (2004) and Pepelko & Withey (1985) are consistent with uncertainty in multiple TK factors giving rise to the wide range of empirical oral to inhalation PoD ratios. Uncertainty is highly increased by not matching for end-points (factor 2 in the above list from Pepelko & Withey, 1985).

In current practice, route-to-route extrapolation for systemic effects is mainly based on comparing assumed maximum absorbed amounts (in percentage of total dose) or absorption rates for the routes under question. Rennen et al. (2004) specifically noted that absorption differences alone are insufficient to account for the range of reported ratios.

Consequently, most of the efforts to extrapolate between routes of exposure have concentrated on development of PBTK models, which can account for the TK factors (e.g. Dallas et al., 1995; Clewell et al., 2001; Sweeney, Saghir & Gargas, 2008; Borghoff, Parkinson & Leavens, 2010; Mielke et al., 2011). The work of Chiu & White (2006) suggests that if these factors are well specified, then the residual uncertainty in route-to-route extrapolation may be small, at least when extrapolating between oral and inhalation routes of exposure. In particular, they noted that route-to-route extrapolation involves only a small number of parameters – many fewer than those required for a full PBTK model – and depends on the internal dose metric associated with the systemic toxic effect. The results of this approach were described in section 4.6.

### A3.5 REFERENCES

Allen BC, Kavlock RJ, Kimmel CA, Faustman EM (1994). Dose–response assessment for developmental toxicity. II. Comparison of generic benchmark dose estimates with no observed adverse effects levels. *Fundam Appl Toxicol.* 23:487–95.

Baird JS, Cohen JT, Graham JD, Shyakhter AI, Evans JS (1996). Noncancer risk assessment: probabilistic characterization of population threshold doses. *J Hum Ecol Risk Assess.* 2(1):79–102.

Batke M, Escher S, Hoffmann-Doerr S, Melber C, Messinger H, Mangelsdorf I (2011). Evaluation of time extrapolation factors based on the database RepDose. *Toxicol Lett.* 205:122–9.

Bokkers BG, Slob W (2005). A comparison of ratio distributions based on the NOAEL and the benchmark approach for subchronic-to-chronic extrapolation. *Toxicol Sci.* 85(2):1033–40.

Bokkers BGH, Slob W (2007). Deriving a data-based interspecies assessment factor using the NOAEL and the benchmark dose approach. *Crit Rev Toxicol.* 37:353–77.

Borghoff SJ, Parkinson H, Leavens TL (2010). Physiologically based pharmacokinetic rat model for methyl tertiary-butyl ether; comparison of selected dose metrics following various MTBE exposure scenarios used for toxicity and carcinogenicity evaluation. *Toxicology.* 275(1–3):79–91.

Chiu WA, White P (2006). Steady-state solutions to PBPK models and their applications to risk assessment. I: Route-to-route extrapolation of volatile chemicals. *Risk Anal.* 26:769–80.

Clewell HJ 3rd, Gentry PR, Gearhart JM, Covington TR, Banton MI, Andersen ME (2001). Development of a physiologically based pharmacokinetic model of isopropanol and its metabolite acetone. *Toxicol Sci.* 63(2):160–72.

Dallas CE, Chen XM, Muralidhara S, Varkonyi P, Tackett RL, Bruckner JV (1995). Physiologically based pharmacokinetic model useful in prediction of the influence of species, dose, and exposure route on perchloroethylene pharmacokinetics. *J Toxicol Environ Health.* 44(3):301–17.

EFSA (2009). Scientific opinion. Use of the benchmark dose approach in risk assessment. Guidance of the Scientific Committee (Question No EFSA-Q-2005-232). *EFSA J.* 1150:1–72.

Freireich EJ, Gehan EA, Rall DP, Schmidt LH, Skipper HE (1966). Quantitative comparison of toxicity of anticancer agents in mouse, rat, dog, monkey and man. *Cancer Chemother Rep.* 50(4):219–44.

Groeneveld C, Hakkert B, Bos P, de Heer C (2004). Extrapolation for exposure duration in oral toxicity: a quantitative analysis of historical toxicity data. *Hum Ecol Risk Assess.* 10:709–16.

Kalberlah F, Schneider K (1998). Quantification of extrapolation factors. Final report of Research Project No. 116 06 113 of the Federal Environmental Agency. Publication Series of the Federal Institute for Occupational Safety and Health, Dortmund. Bremerhaven: Wirtschaftsverlag NW (Research Report Fb 797).

Kalberlah F, Föst U, Schneider K (2002). Time extrapolation and interspecies extrapolation for locally acting substances in case of limited toxicological data. *Ann Occup Hyg.* 46:175–85.

Kramer HJ, Van den Ham WA, Slob W, Pieters MN (1996). Conversion factors estimating indicative chronic no-observed-adverse-effect levels from short-term toxicity data. *Regul Toxicol Pharmacol.* 23(3):249–55.

McNamara BP (1976). Concepts in health evaluation of commercial and industrial chemicals. In: Mehlman MA, Shapiro RE, Blumenthal H, editors. *Advances in modern toxicology*. Vol. 1. New concepts in safety evaluation. Washington (DC): Hemisphere.

Mielke H, Abraham K, Gotz M, Vieth B, Lampen A, Luch A et al. (2011). Physiologically based toxicokinetic modelling as a tool to assess target organ toxicity in route-to-route extrapolation – The case of coumarin. *Toxicol Lett*. 202(2):100–10.

Nessel CS, Lewis SC, Stauber KL, Adgate JL (1995). Subchronic to chronic exposure extrapolation: toxicologic evidence for a reduced uncertainty factor. *Hum Ecol Risk Assess*. 1(5):516–26.

Pepelko WE, Withey JR (1985). Methods for route-to-route extrapolation of dose. *Toxicol Ind Health*. 1(4):153–70.

Pieters MN, Kramer HJ, Slob W (1998). Evaluation of the uncertainty factor for subchronic-to-chronic extrapolation: statistical analysis of toxicity data. *Regul Toxicol Pharmacol*. 27(2):108–11.

Price PS, Keenan RE, Swartout JC (2008). Characterizing interspecies uncertainty using data from studies of anti-neoplastic agents in animals and humans. *Toxicol Appl Pharmacol*. 233(1):64–70.

Rennen MAJ, Hakkert BC, Stevenson H, Bos PMJ (2001). Data-base derived values for the interspecies extrapolation. *Comments Toxicol*. 7:423–36.

Rennen MA, Bouwman T, Wilschut A, Bessems JG, Heer CD (2004). Oral-to-inhalation route extrapolation in occupational health risk assessment: a critical assessment. *Regul Toxicol Pharmacol*. 39:5–11.

Rulis AM, Hattan DG (1985). FDA's priority-based assessment of food additives. II. General toxicity parameters. *Regul Toxicol Pharmacol*. 5:152–74.

Sweeney LM, Saghir SA, Gargas ML (2008). Physiologically based pharmacokinetic model development and simulations for ethylene dichloride (1,2-dichloroethane) in rats. *Regul Toxicol Pharmacol*. 51(3):311–23.

Vermeire T, Stevenson H, Pieters MN, Rennen M, Slob W, Hakkert BC (1999). Assessment factors for human health risk assessment: a discussion paper. *Crit Rev Toxicol*. 29:439–90.

Weil CS, McCollister DD (1963). Relationship between short- and long-term feeding studies in designing an effective toxicity test. *Agric Food Chem*. 11:486–91.

## ANNEX 4: DETAILED RESULTS OF ANALYSES OF HISTORICAL DATA ON INTRASPECIES HUMAN VARIABILITY

### A4.1 Introduction

As discussed in section 3.2, one individual is considered to be X times more sensitive than another if he or she experiences a biological response of the same magnitude (or severity) at an X-fold lower dose. In other words, the equipotent dose is X-fold lower. The variation in sensitivity in a population is typically characterized as a lognormal distribution of equipotent doses. The median of that distribution is 1, representing the typical individual. Therefore, the distribution is fully characterized by a single parameter (i.e. the variability parameter). The measure that will be used here is the  $\log(\text{GSD}_H)$  – the base 10 logarithm of the geometric standard deviation (GSD) for human (subscript H) variability – or, equivalently, the standard deviation of the log-transformed individual equipotent doses.<sup>22</sup> Table A4.1 provides a convenient translation between lognormal variability expressed as  $\log(\text{GSD}_H)$  and other measures that may be more familiar to some readers.

**Table A4.1: Translation guide for understanding lognormal variability expressed as  $\log(\text{GSD}_H)$**

$\log(\text{GSD}_H)$	$\text{GSD}_H$	Coefficient of variation (%)	P95/P50 of equipotent HDs
0.05	1.12	11.6	1.21-fold
0.1	1.26	23.3	1.46-fold
0.15	1.41	35.6	1.76-fold
0.2	1.58	48.6	2.13-fold
0.25	1.78	62.7	2.58-fold
0.3	2.00	78.2	3.11-fold
0.4	2.5	116	4.55-fold
0.5	3.2	166	6.64-fold
0.6	4	240	9.70-fold
0.7	5	353	14.2-fold
0.8	6.3	536	20.7-fold

$\text{GSD}_H$ : geometric standard deviation for interindividual variability in the human equipotent dose distribution; HD: human dose; P50: 50th percentile; P95: 95th percentile

The human equipotent dose distribution, and the uncertainty in that distribution, may be informed by collecting intraspecies variation estimates for various chemicals. The relevant studies that did so are reviewed below. The results from these studies indicated that the individual chemical  $\log(\text{GSD}_H)$  values can be closely described by lognormal distributions themselves (Hattis & Lynch, 2007). The distinction between the variability distribution for equipotent doses and the associated uncertainty distribution for the intraspecies  $\log(\text{GSD}_H)$  is further illustrated in Fig. 4.1 of the main text. It should be noted that both distributions are assumed to be lognormal and that care needs to be taken not to confuse the two. It should also

<sup>22</sup> A lognormal distribution for the equipotent doses is equivalent to a normal distribution for the log-transformed equipotent doses. The variation may be given in terms of the SD of the normal distribution of log doses or by the back-transformation of that, which is called the GSD ( $= 10^{\text{SD}}$ ). Therefore,  $\log(\text{GSD}) = \text{SD}$  of the log doses. To avoid confusion with the GSD for uncertainty distributions, this specific GSD for human variability is given a subscript “H”.

be noted that the notation  $\log(\text{GSD}_U)$  is used for the uncertainty in  $\log(\text{GSD}_H)$ , the interindividual variation in equipotent doses.

As discussed in section 4.5, intraspecies variability is further split up into two subcomponents:

1. TK variability is defined as differences among people in the external dose required to produce a similar systemic internal dose (concentration–time combination for systemically acting agents), usually measured in the blood.
2. TD variability is defined as differences among people in the internal dose required to produce an effect of defined degree or severity ( $M$ ).

Below, intraspecies distributions are suggested based on a review of historical data from relevant studies. The presentation here draws on two substantial compilations of human variability information, one by Hattis and colleagues (Hattis et al., 2001; Hattis, Baird & Goble, 2002; Hattis & Lynch, 2007; Hattis, 2013) and the other by Renwick, Dorne and colleagues (Renwick & Lazarus, 1998; Dorne, Walton & Renwick, 2005). These compilations are largely, but not solely, assembled from measurements made for pharmaceuticals.

#### A4.2 Observed intraspecies variation in toxicokinetics based on historical data

The most commonly used measures of internal dose are the area under the curve for the plasma or blood concentration against time per unit of external dose (AUC) and the maximum concentration in blood or plasma achieved over time ( $C_{\max}$ ). Although the concentration at the active site or target tissue can also be a measure of internal dose, it is less commonly available due to data limitations. For this reason, it is conventional to split TK and TD at circulating concentrations of the parent compound or active metabolite (Meek & Renwick, 2006), with delivery to or concentration at the active site being considered part of TD (see below).

An analysis of oral AUC variability data for 31 chemicals compiled by Renwick & Lazarus (1998) yielded a geometric mean for  $\log(\text{GSD}_H)$  of 0.15 and a P95/P50 ratio of 2.1. The most recent compilation of observed variability in oral AUC values is by Hattis & Lynch (2007), shown in Tables A4.2 and A4.3. The 31 data sets from adult populations yield a GM for the  $\log(\text{GSD}_H)$  of 0.16 and a GSD of 1.71, implying a P95/P50 ratio of 2.42, slightly higher than the previous analysis. The six data sets that include children under 12 years of age are shown in Table A4.3, which yield a GM of 0.20 and GSD of 1.76.

The results for adults only and those including children are not statistically significantly different (using the  $F$ -test for the variances and the  $t$ -test for the means, not shown). Therefore, it is not unreasonable to combine these data sets, with the following results:

TK variability (37 oral AUCs):

$$\text{GM of } \log(\text{GSD}_H)_{\text{TK}} = 0.167, \text{ P95/P50} = 2.43 \text{ [(P05, P95) = (0.0687, 0.407)]}$$

**Table A4.2: Listing of individual observations of  $\log(\text{GSD}_H)$  for adult toxicokinetics, as reflected in variability in systemic concentrations (AUC) after oral dosing**

<i>Agent</i>	<i>Log(GSD<sub>H</sub>)</i>	<i>N</i>
Ampicillin	0.070	5
Aspirin	0.097	10
Bromazepam	0.134	10
Brotizolam	0.110	8
Cimetidine	0.147	20
Clozapine	0.307	7
Clozapine- <i>N</i> -oxide	0.206	6
Cyclosporin	0.117	22
Dapsone	0.098	7
Dixyrazine	0.403	8
Enalapril	0.132	18
Enalaprilat (active metabolite of enalapril)	0.151	18
Flunisolide	0.269	12
Indomethacin	0.269	5
Ketoprofen – active <i>S</i> -enantiomer	0.127	14
Lorazepam	0.123	6
Metoclopramide	0.196	6
Mexiletine	0.070	5
Moxonidine	0.170	8
Norclozapine	0.451	7
Nortriptyline	0.141	6
Phenylpropanolamine	0.109	7
Praziquantel	0.406	8
Prifinium bromide	0.109	6
Sulfamethazine	0.210	5
Timolol	0.278	5
Treosulfan	0.170	10
Triazolam	0.306	8
Trimethoprim	0.104	6
Valproic acid	0.056	10
Viloxazine	0.211	16
<b>GM of <math>\log(\text{GSD}_H)_{TK}</math></b>	<b>0.161</b>	
<b><math>\text{GSD}_U</math> of <math>\log(\text{GSD}_H)_{TK}</math></b>	<b>1.711</b>	

AUC: area under the concentration–time curve; GM: geometric mean;  $\text{GSD}_H$ : geometric standard deviation for interindividual variability in the human equipotent dose distribution;  $\text{GSD}_U$ : a measure of the uncertainty in the  $\text{GSD}_H$ ; *N*: number of adults; TK: toxicokinetics; U: uncertainty

This distribution can be used as a preliminary uncertainty distribution, given that in the most common case of chronic end-points mediated systemically, AUCs are the most commonly accepted TK measure.

**Table A4.3: Listing of additional individual observations of  $\log(\text{GSD}_H)$  for toxicokinetics, as reflected in variability in systemic concentrations (AUC) after oral dosing in groups that include children under the age of 12**

<i>Agent</i>	<i>Log(GSD<sub>H</sub>)</i>	<i>N</i>
Captopril	0.338	8
Carbamazepine	0.198	39
Metoclopramide	0.348	9
Metronidazole	0.147	13
Nifedipine	0.265	9
Tobramycin	0.080	7
<b>GM of <math>\log(\text{GSD}_H)_{TK}</math></b>	<b>0.204</b>	
<b>GSD<sub>U</sub> of <math>\log(\text{GSD}_H)_{TK}</math></b>	<b>1.760</b>	

AUC: area under the concentration–time curve; GM: geometric mean; GSD<sub>H</sub>: geometric standard deviation for interindividual variability in the human equipotent dose distribution; GSD<sub>U</sub>: a measure of the uncertainty in the GSD<sub>H</sub>; N: number of individuals; TK: toxicokinetics

Alternatively, from the same database, a similar analysis of human interindividual variability in the 29 oral  $C_{max}$  values (not shown), five of which contain data on children under the age of 12, yields a GM of 0.155 and a P95/P50 ratio of 2.90. However, the results for AUCs and  $C_{max}$  values are not statistically significantly different (by the *t*-test for the means and the *F*-test for the variances, not shown), so combining these data is also an option, yielding a GM of 0.162 and a P95/P50 ratio of 2.62. Note that these alternative options lead to results very similar to those of the proposed preliminary distribution above.

#### A4.3 Observed intraspecies variation in toxicodynamics based on historical data

Measures of interindividual variability in TD for systemic toxicants are more diverse and require more elaborate analysis than is needed for the measures of TK variability summarized above. Hattis et al. (1999) subdivided measured parameters with information on TD variability into three causal steps:

1. TD1: Variability in the effective delivery of systemically available chemical to the site of action;
2. TD2: Variability in the concentration of the chemical at the active site needed to change a physiological parameter by a specific amount (e.g. a particular change in a measure of blood clotting tendency; 50% of maximal change in a measure of brain electrical activity); and
3. TD3: Variability in functional reserve capacity – the amount of physiological parameter change needed to elicit a specific quantal biological response at a given degree of severity (e.g. headache, eye irritation, nausea, overtly dose-limiting toxicity for anti-cancer agents).

Table A4.4 reports summary observations of variability among chemicals in parameters that relate to various TD steps. It can be seen in Table A4.4 that observations of local (portal of entry) responses or related parameter changes (first three rows) give rise to much larger GM values for observed interindividual variability compared with observations of non-immune-related systemic effects (last three rows). This may be due, in part, to the fact that there is

substantially more variability in TD<sub>1</sub> for localized effects at the portal of entry compared with systemic effects. Additionally, it should be noted that many of the local effects are immune system mediated, and such responses tend to be more variable than responses mediated by other processes. This is consistent with the observation that interindividual variability in immune-related systemic effects is similar to that for portal effects.

**Table A4.4: Distributions of log(GSD<sub>H</sub>) among chemicals for various toxicodynamic steps**

<i>GM</i>	<i>P95/P50</i>	<i>N</i>	<i>Type of response parameter</i>	<i>TK and TD steps involved</i>
0.469	2.20	9	Continuous inhalation parameter change (e.g. FEV <sub>1</sub> change) <sup>a</sup>	TD <sub>1</sub> + TD <sub>2</sub>
0.550	3.05	7	Quantal responses to inhalation (e.g. wheeze, throat irritation) <sup>a</sup>	TD <sub>1</sub> + TD <sub>2</sub> + TD <sub>3</sub>
0.544	2.37	5	Skin hypersensitivity and irritation responses <sup>a</sup>	TD <sub>1</sub> + TD <sub>2</sub> + TD <sub>3</sub>
0.536	1.86	4	Internal concentrations producing specific immune-related physiological parameter changes <sup>b,c</sup>	TD <sub>1</sub> + TD <sub>2</sub>
0.195	2.76	18	Internal concentrations producing specific non-immune-related physiological parameter changes <sup>b,c</sup>	TD <sub>1</sub> + TD <sub>2</sub>
0.256	2.89	16	Internal concentrations producing non-immune-related quantal responses <sup>b,d,e</sup>	TD <sub>1</sub> + TD <sub>2</sub> + TD <sub>3</sub>
0.242	4.27	10	Non-immune-related quantal response in relation to external exposures <sup>b,d</sup>	TK + TD <sub>1</sub> + TD <sub>2</sub> + TD <sub>3</sub>

FEV<sub>1</sub>: forced expiratory volume in 1 second; GM: geometric mean; N: number of agents studied; P50: 50th percentile; P95: 95th percentile; TD: toxicodynamics; TK: toxicokinetics

<sup>a</sup> Portal of entry response.

<sup>b</sup> Effects mediated systemically.

<sup>c</sup> The GMs for non-immune-related and immune-related continuous responses were statistically significantly different.

<sup>d</sup> Insufficient data on immune-related effects for subgroup analysis.

<sup>e</sup> Includes effects of varying severity; however, there was no statistically significant difference between mild or moderate effects and more severe effects, so the data were pooled.

Source: Hattis & Lynch (2007)

The diversity of TD end-points complicates the selection of a preliminary uncertainty distribution for intraspecies TD. Because the most common cases encountered in hazard characterization involve non-immune-related effects mediated systemically, rows 5 and 6 of Table A4.4 were chosen as the basis for deriving a preliminary distribution. The underlying data are shown in Tables A4.5 and A4.6. Additionally, these distributions are not statistically significantly different (analysis not shown), so for the purposes of deriving a preliminary distribution, the data were combined, yielding the following:

TD variability:

$$GM \text{ of } \log(GSD_H)_{TD} = 0.221, P95/P50 = 2.85 [(P05, P95) = (0.0776, 0.631)]$$

**Table A4.5: Listing of individual observations of  $\log(\text{GSD}_H)$  for continuous physiological parameter changes in relation to internal measures of systemic exposure**

<i>Agent</i>	<i>Parameter measured</i>	<i>Log(GSD<sub>H</sub>)</i>	<i>N</i>
Alfentanil	EC <sub>50</sub> : Effect site concentration producing 50% of predetermined maximal EEG changes	0.214	5
Befloxatone	EC <sub>50</sub> : Effect site concentration producing 50% of the maximal extent of monoamine oxidase-A inhibition (as characterized by the decrease in 3,4-dihydroxyphenylglycol) as attributed to befloxadone	0.194	12
Benazepril	IC <sub>50</sub> : Plasma concentration producing 50% inhibition of angiotensin converting enzyme	0.145	16
Doxazosin	Reduction in diastolic blood pressure per unit drug concentration at effect site	0.208	10
Doxazosin	Reduction in systolic blood pressure per unit drug concentration at effect site	0.127	10
Enalaprilat (active metabolite of enalapril)	IC <sub>50</sub> : Plasma concentration producing 50% inhibition of angiotensin converting enzyme	0.173	12
Enalaprilat (active metabolite of enalapril)	IC <sub>50</sub> : Plasma concentration producing 50% inhibition of angiotensin converting enzyme	0.156	15
Enalaprilat (active metabolite of enalapril)	IC <sub>50</sub> : Plasma concentration producing 50% inhibition of angiotensin converting enzyme	0.251	18
Fentanyl	EC <sub>50</sub> : Measured as the effect site concentration producing 50% of a predetermined maximal EEG change in 3–10 minutes	0.302	5
Furosemide	Diuretic efficiency (mL/μg) (drug-induced urine flow/drug excretion rate)	0.048	8
Furosemide	Natriuretic efficiency (mL/μg) (drug-induced response/drug excretion rate)	0.066	8
Imiprimine	Proportion of patients receiving more than 95% of their individual maximal response in relation to plasma concentration	0.253	15
Ketoprofen – active S-enantiomer	EC <sub>50</sub> : Unbound S-enantiomer concentration required for 50% inhibition of platelet thromboxane A2 generation during controlled clotting of whole blood	0.293	14
Levodopa	EC <sub>50</sub> : Concentration of drug producing half-maximal anti-parkinsonism effect	0.300	37
MK852	EC <sub>50</sub> : Concentration in vitro to achieve 50% inhibition of ADP-induced platelet aggregation	0.171	7
Oxypurinol (active metabolite of allopurinol)	EC <sub>50</sub> : Concentration of oxypurinol in plasma to achieve 50% inhibition of the ratio of 1-methyluric acid to 1-methylxanthine in urine (a measure of xanthine oxidase inhibition)	0.160	5
Sotalol (d-)	EC <sub>50</sub> : Prolongation of cardiac Q-Tc interval (antiarrhythmia agent)	0.802	24

Table A4.5 (continued)

<i>Agent</i>	<i>Parameter measured</i>	<i>Log(GSD<sub>H</sub>)</i>	<i>N</i>
Trefentanil	EC <sub>50</sub> : Measured as the effect site concentration producing 50% of a predetermined maximal EEG change in 3–10 minutes	0.319	7
<b>GM of log(GSD<sub>H</sub>)<sub>TD</sub></b>		<b>0.195</b>	
<b>GSD<sub>U</sub> of log(GSD<sub>H</sub>)<sub>TD</sub></b>		<b>1.856</b>	

ADP: adenosine diphosphate; EC<sub>50</sub>: median effective concentration; EEG: electroencephalographic; GM: geometric mean; GSD<sub>H</sub>: geometric standard deviation for interindividual variability in the human equipotent dose distribution; GSD<sub>U</sub>: a measure of the uncertainty in the GSD<sub>H</sub>; IC<sub>50</sub>: median inhibitory concentration; TD: toxicodynamics

Source: Hattis & Lynch (2007)

Table A4.6: Listing of individual observations of log(GSD<sub>H</sub>) for quantal biological responses in relation to internal measures of systemic exposure

<i>Agent</i>	<i>Parameter measured</i>	<i>Log(GSD<sub>H</sub>)</i>	<i>Population studied</i>
Cadmium	High (over the 2.5th percentile of an unexposed population) $\beta_2$ -microglobulin urinary excretion compared with urinary cadmium	0.360	3115 residents of contaminated community 50+ years of age
Cadmium	High (over the 2.5th percentile of an unexposed population) $\beta_2$ -microglobulin urinary excretion compared with a dose metric of cumulative cadmium blood concentration $\times$ time	0.556	437 workers exposed via inhalation
Carbamate and phosphate anticholinesterases (unspecified)	Deaths/red blood cell cholinesterase inhibition "hits"	0.145	20 attempted suicide cases
Digoxin	Digoxin toxicity in relation to serum digoxin concentration	0.133	710 patients requiring digoxin therapy for heart failure or atrial fibrillation with tachycardia – half inpatients, half outpatients, ages 51–81+ years
Haloperidol (antipsychotic)	Haloperidol toxicity (minimum of four other signs plus, in some cases, seizures, catatonia and mental confusion) in relation to maximum blood level	0.115	43 patients with chronic schizophrenia
Methylmercury	Deaths/blood level	0.128	125 Iraqi adults who consumed contaminated food
Methylmercury	Hearing defects/blood level	0.143	125 Iraqi adults who consumed contaminated food
Methylmercury	Dysarthria/blood level	0.186	125 Iraqi adults who consumed contaminated food
Methylmercury	Ataxia/blood level	0.232	125 Iraqi adults who consumed contaminated food

Table A4.6 (continued)

<i>Agent</i>	<i>Parameter measured</i>	<i>Log(GSD<sub>H</sub>)</i>	<i>Population studied</i>
Methylmercury	Paraesthesia/blood level	0.382	125 Iraqi adults who consumed contaminated food
Methylmercury	Visual effects/blood level	0.458	125 Iraqi adults who consumed contaminated food
Military anticholinesterase agents	Psychomotor depression/blood cholinesterase	0.232	93 military test subjects
Military anticholinesterase agents	Anxiety/blood cholinesterase	0.475	93 military test subjects
Military anticholinesterase agents	Unusual dreams/blood cholinesterase	0.815	93 military test subjects
Tenecteplase	Achievement of a specific degree of cardiac blood flow (unblocking of a clot) following an infarction in relation to the 2–90 minute AUC of a tissue plasminogen activator	0.123	85 patients 18–80 years of age with acute myocardial infarction within 12 hours of symptom onset, requiring thrombolytic therapy
Trinitrotoluene	Cataracts in relation to trinitrotoluene–haemoglobin adducts	0.502	117 workers (China)
<b>GM of log(GSD<sub>H</sub>)<sub>TD</sub></b>			<b>0.256</b>
<b>GSD<sub>U</sub> of log(GSD<sub>H</sub>)<sub>TD</sub></b>			<b>1.891</b>

AUC: area under the concentration–time curve; GM: geometric mean; GSD<sub>H</sub>: geometric standard deviation for interindividual variability in the human equipotent dose distribution; GSD<sub>U</sub>: a measure of the uncertainty in the GSD<sub>H</sub>; TD: toxicodynamics

Source: Hattis & Lynch (2007)

For portal and immune-related effects, this distribution might underestimate the degree of intraspecies variability.

The analytical results shown at the bottom of Tables A4.5 and A4.6 are based on simple unweighted averages of the logs of the observed log(GSD<sub>H</sub>)<sub>TD</sub>. This results from the assumption that the differences among the observations are likely to primarily reflect real differences among chemicals and types of responses and that there should therefore be no differential weighting of the observations by the size of the populations studied or other measures of the statistical strength within each data set. In this view, the GSD<sub>U</sub> results are seen as reflecting real uncertainties faced by an analyst or risk manager for an agent with non-immune-mediated systemic effects that is essentially a random draw from those that have been previously studied.

Prior analyses by Hattis and colleagues (Hattis, Baird & Goble, 2002; Hattis & Lynch, 2007) have used a more complex “variance allocation modelling” approach that produces somewhat different and more diverse estimates of log(GSD<sub>H</sub>)<sub>TD</sub> for different types of measurement end-points and for effects of different severities. The basic idea is to treat each observation as the result of a combination of the variances from specific components of the causal chain reflected in each measurement – that is, any TK steps and the TD steps defined at the

beginning of section A4.3 above. Thus, each of the observations of Table A4.5 is modelled as:

$$\text{Log(GSD)} = \sqrt{\text{TD}_1\text{GSD}^2 + \text{TD}_2\text{GSD}^2} \quad (\text{A4-1})$$

whereas each of the observations of Table A4.6 is modelled as:

$$\text{Log(GSD)} = \sqrt{\text{TD}_1\text{GSD}^2 + \text{TD}_2\text{GSD}^2 + \text{TD}_3\text{GSD}^2} \quad (\text{A4-2})$$

where each of the variance components inside the square root symbols is regarded as a variable whose value is estimated by minimizing the weighted sum of the squares of the observed versus expected log(GSD) observations. Additional observations include TK components. Moreover, separate TD<sub>2</sub> and TD<sub>3</sub> parameters are used in the model for immune versus non-immune parameter changes and responses; for TD<sub>3</sub>, responses are roughly classified into three categories by severity.<sup>23</sup> The weights given each observation are the reciprocals of the measurement variance for each data point, estimated as described in Hattis & Lynch (2007).

Applying this model to the current data set of log(GSD) observations, TK measurements, physiological parameter changes and effects following systemic exposure yields the results reported in Table A4.7.

**Table A4.7: Central estimates of the GSD<sub>H</sub> for different components of TD variability, alone and in combination**

<b>Component of TD variability (and combinations)</b>	<b>Central estimate of the GSD<sub>H</sub></b>
TD <sub>1</sub>	0.142
TD <sub>2</sub>	0.180
TD <sub>3</sub> (mild effects)	0.231
TD <sub>3</sub> (moderate or severe effects)	Very near zero
TD <sub>1</sub> + TD <sub>2</sub> (physiological parameter changes or moderate or severe effects in relation to systemic exposure)	0.230
TD <sub>1</sub> + TD <sub>2</sub> + TD <sub>3</sub> (mild effects in relation to systemic exposure base)	0.326
TK + TD <sub>1</sub> + TD <sub>2</sub> (physiological changes or moderate or severe effects in relation to external dose)	0.281
TK + TD <sub>1</sub> + TD <sub>2</sub> + TD <sub>3</sub> (mild effects in relation to external dose)	0.363

GSD<sub>H</sub>: geometric standard deviation for interindividual variability in the human equipotent dose distribution; TD: toxicodynamics; TK: toxicokinetics

The central result of 0.230 for TD<sub>1</sub> + TD<sub>2</sub> from this approach is therefore slightly higher than the direct unweighted calculation of 0.195 in Table A4.5. Similarly, the TD<sub>1</sub> + TD<sub>2</sub> + TD<sub>3</sub> result for mild effects of 0.326 implies greater variability than the bottom line derivation of 0.256 from the bottom line of Table A4.6. Some of this latter difference may have resulted from measurement errors in the very limited number of “mild effect” observations; however, some may be attributable to real differences in human variability for the more severe effects. It would make biological sense for homeostatic defence mechanisms against really severe

<sup>23</sup> Mild (e.g. analgesia from dental pain, paraesthesia from methylmercury, liver dysfunction indicated by elevated levels of serum aminotransferase, nausea in relation to ingestion of dissolved copper), moderate (high β<sub>2</sub>-microglobulin urinary excretion compared with urinary cadmium; vomiting in relation to dissolved copper) and severe (death or ataxia from methylmercury).

outcomes to be both more vigorous and less variable among people than defences against less severe responses that would not be expected to compromise evolutionary “fitness”. We expose these differences in the outcome of various analytical approaches to help both assessors and managers consider what technical and management responses are appropriate in the light of these different kinds of results and what additional research might be desirable in order to achieve better quantification of variability in susceptibility for effects of different severity for different kinds of toxicants.

Other things being equal, it would be reasonable to couple these central estimates with similar estimates of P95/P05 uncertainty ratios as are derived in the section below for the simpler unweighted variability estimates.

#### A4.4 Estimated uncertainty in overall intraspecies variability

The overall intraspecies distribution of (external) equipotent doses can be estimated by combining the separate distributions for TK and TD assuming independence, as follows:

$$[\log(\text{GSD}_{\text{H-overall}})]^2 = [\log(\text{GSD}_{\text{H-TK}})]^2 + [\log(\text{GSD}_{\text{H-TD}})]^2 \quad (\text{A4-3})$$

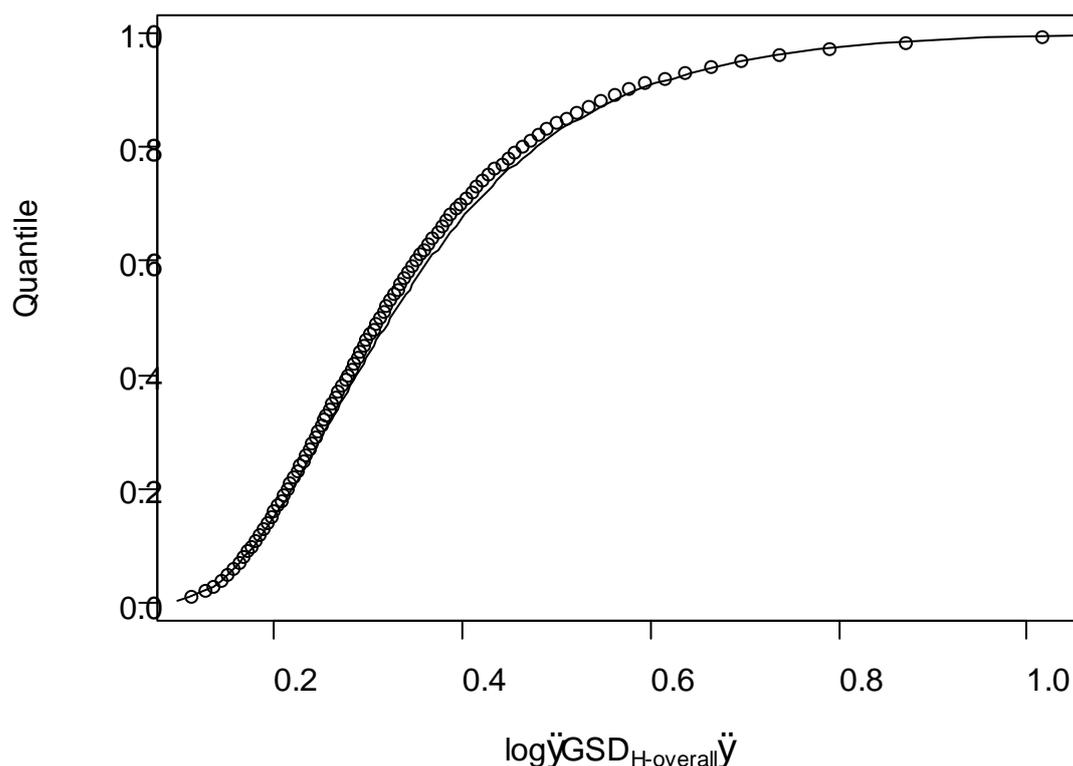
Because  $\log(\text{GSD}_{\text{H-TK}})$  and  $\log(\text{GSD}_{\text{H-TD}})$  are each assumed to be lognormally distributed, there is no exact form for the distribution of  $\log(\text{GSD}_{\text{H-overall}})$ . It is therefore approximated using a lognormal distribution with the same P05 and P95, as evaluated using Monte Carlo simulation. Specifically, values of  $\log(\text{GSD}_{\text{H-TK}})$  and  $\log(\text{GSD}_{\text{H-TD}})$  are drawn randomly from their respective lognormal distributions, resulting in a collection of random estimates of  $\log(\text{GSD}_{\text{H-overall}})$ . A lognormal distribution is then fit with the same P05 and P95 as for the Monte Carlo samples. The result of this simulation using  $10^7$  random samples is:

Intraspecies variability:

$$\text{GM of } \log(\text{GSD}_{\text{H}}) = 0.324, \text{ P95/P50} = 2.152 \text{ [(P05, P95) = (0.151, 0.697)]}$$

A comparison of the empirical combined distribution for  $10^7$  Monte Carlo samples and the lognormal approximation is shown in Fig. A4.1.

It is also useful to compare this derived distribution with the available data on overall intraspecies variability, discussed in section A.3. Specifically, the last row of Table A4.4 summarizes the available data on variability in external exposure causing quantal effects. The 10 available data sets are shown in Table A4.8, along with summary statistics. Although the GM and P95/P50 appear to be somewhat different from the values derived for the distribution above, the two distributions are not statistically significantly different (by Kolmogorov-Smirnov test, not shown).



**Fig. A4.1: Uncertainty distribution of intraspecies variability estimated using historical data.** The circles are the 1–99% quantiles based on  $10^7$  Monte Carlo samples assuming that  $\log(\text{GSD}_{\text{H-TK}})$  and  $\log(\text{GSD}_{\text{H-TD}})$  are each lognormally distributed with GM and P95/P50 as specified in sections A4.2 and A4.3. The line represents quantiles of the lognormal distribution with the same P05 and P95 as for the Monte Carlo sample distribution.

**Table A4.8: Listing of individual observations of  $\log(\text{GSD}_{\text{H}})$  for quantal biological responses in relation to external exposure.**

<i>Agent</i>	<i>Parameter measured</i>	<i>Log(GSD<sub>H</sub>)</i>	<i>Population studied</i>
Copper	Nausea and vomiting in relation to oral drinking-water concentrations	0.293	61 healthy adults, aged 18–50 years, 31 women and 30 men
Cyclophosphamide	Liver dysfunction (elevated levels of serum aminotransferase) in relation to intravenous dose	0.269	Cancer patients (leukaemias and other blood cell malignancies), aged 12–72 years, mean 54 years
Ibuprofen	Analgesia from dental pain (not taking medication at 3 and 6 hours after procedure)	0.546	304 patients undergoing surgical removal of molars
Isoflurane	End tidal concentration for anaesthesia (not moving in response to stimulus)	0.070	36 premature infants
Midazolam	“Adequate” sedation/drowsiness	0.508	85 paediatric patients (aged 0.9–15.7 years, mean age 6.6 years, 51 males) requiring one-time sedation before surgical or non-surgical procedures
Pyrazoloacridine	Neutropenia (two levels)	0.266	20 patients with advanced cancer, undergoing chemotherapy

Table A4.8 (continued)

<i>Agent</i>	<i>Parameter measured</i>	<i>Log(GSD<sub>H</sub>)</i>	<i>Population studied</i>
Rocuronium	Creation of conditions for intubation (two levels – “excellent” and “good”)	0.328	94 adult patients
Sevoflurane	End tidal concentration for anaesthesia (not moving in response to stimulus)	0.037	20 children 3–5 years old
Suramin	Dose-limiting toxicity, including malaise, neurotoxicity, pericardial effusion and coagulopathy	0.497	Patients with advanced cancer, undergoing chemotherapy
Suxamethonium	Suppression of coughing (two levels) on intubation	0.284	60 adult patients 17–49 years of age requiring intubation for oral surgery
<b>GM of log(GSD<sub>H-overall</sub>)</b>			<b>0.242</b>
<b>GSD<sub>U</sub> of log(GSD<sub>H-overall</sub>)</b>			<b>2.418</b>

GM: geometric mean; GSD<sub>H</sub>: geometric standard deviation for interindividual variability in the human equipotent dose distribution

Source: Hattis & Lynch (2007)

#### A4.5 References

Dorne JL, Walton K, Renwick AG (2005). Human variability in xenobiotic metabolism and pathway-related uncertainty factors for chemical risk assessment: a review. *Food Chem Toxicol.* 43:203–16.

Hattis D (2013). Human interindividual variability in parameters related to susceptibility for toxic effects [website]. Worcester (MA): Clark University (<http://www2.clarku.edu/faculty/dhattis>, accessed 30 April 2014).

Hattis D, Lynch MK (2007). Empirically observed distributions of pharmacokinetic and pharmacodynamic variability in humans – Implications for the derivation of single point component uncertainty factors providing equivalent protection as existing RfDs. In: Lipscomb JC, Ohanian EV, editors. *Toxicokinetics in risk assessment*. Informa Healthcare USA, Inc.; 69–93.

Hattis D, Baird S, Goble R (2002). A straw man proposal for a quantitative definition of the RfD. *Drug Chem Toxicol.* 25:40–36.

Hattis D, Banati P, Goble R, Burmaster DE (1999). Human interindividual variability in parameters related to health risks. *Risk Anal.* 19:705–20.

Hattis D, Russ A, Goble R, Banati P, Chu M (2001). Human interindividual variability in susceptibility to airborne particles. *Risk Anal.* 21:585–99.

Meek B, Renwick A (2006). Guidance for the development of chemical-specific adjustment factors: integration with mode of action frameworks. In: Lipscomb JC, Ohanian EV, editors. *Toxicokinetics and risk assessment*. New York (NY): Informa Healthcare; 27–46.

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

## ANNEX 5: CASE-STUDY: DEOXYNIVALENOL (DON)

### A5.1 Introduction

#### A5.1.1 General note

The following example draws upon a risk characterization that was performed by the Dutch National Institute for Public Health and the Environment (Rijkinstituut voor Volksgezondheid en Milieu, or RIVM) (see Pieters et al., 2001; Pieters, Bakker & Slob, 2004; Bokkers et al., 2009). For the sake of simplicity, the choices of studies as well as of adversity thresholds have been kept here as they were made by RIVM at that time. In some instances, a reader familiar with the toxicity database of DON might feel that he or she might have chosen differently. The same might hold true for other choices made by the authors of the present text, such as the percentiles of the target population to be protected or values for certain default assessment factors.

It is emphasized that this example by no means aims to provide a definitive hazard assessment for DON or to be prescriptive in terms of the choices referred to above. Its sole purpose is to provide a practical exercise of applying the principles of uncertainty analysis as presented in this guidance document.

Before starting with this example, the reader should be familiar with the following aspects of the framework, as explained in the main text of this guidance document:

- the concept of the target human dose,  $HD_M^I$ , associated with a particular magnitude of effect  $M$  at a particular population incidence  $I$  (sections 2.3, 2.4.1 and 3.2);
- the meaning of the terms “coverage” and “degree of uncertainty” (section 2.4.2);
- the APROBA tool (section 3.3.2); and
- the concept of generic primary uncertainty distributions (section 4).

#### A5.1.2 Problem formulation

DON is a mycotoxin found in *Fusarium*-infected cereals; hence, humans may be exposed to DON when they consume cereal-based foodstuffs. The presence of DON in cereals cannot be prevented, but if appropriate countermeasures are taken, such as storing cereals under non-humid conditions, the levels can be reduced. The risk assessment question is whether current exposure to DON may be regarded as acceptable with respect to potential health effects in the human population. To that end, this hazard characterization aims to derive a health-based guidance value, in this case an RfD, for potentially critical effects, which will then be compared with the exposure estimates for the target populations.

#### A5.1.3 General approach

As discussed in the main text, hazard characterization may be done by a tiered approach, where, after each tier, the question is posed: “Do we know enough, or is a higher-tier assessment needed?” The question “Do we know enough?” may be answered by evaluating the “overall” uncertainty in the hazard assessment outcome together with information on human exposure. Information on expected exposure was taken from the above RIVM publications and is not further discussed here, as this case-study illustrates only the principle

of evaluating uncertainties in hazard characterization. The uncertainties in each tier of assessment will be probabilistically evaluated in a generic way using the tool APROBA.

Several studies have shown adverse effects after DON administration to experimental animals – for example, on body weight, prenatal development of offspring and fertility. When starting hazard assessment in such a situation, it is most often not clear which of these effects will ultimately be the basis of an RfD. Therefore, a recommendable procedure would be to perform the assessment in parallel for all studies/end-points for which a hazard has been identified and to conclude on the basis of the various RfDs at the end.

For ease of understanding, however, this case-study first addresses hazard and uncertainty assessment for one end-point only (sections A5.2 and A5.3), as the focus of this annex is on illustrating the general approach. After that, inclusion of further end-points is discussed (section A5.4).

As a first tier, a “traditional” hazard assessment using the NOAEL approach with default assessment factors is performed in order to derive a “deterministic” (non-probabilistic) RfD for the human target population (section A5.2.1).

Then (in section A5.2.2), for each aspect of the hazard characterization, the uncertainty is characterized by the corresponding generic uncertainty distribution given in Table 4.6 of the main document. Where substance- or study-specific data are available, the parameters underlying the generic distributions are replaced by parameters (and, hence, distributions) that are more adequate for the specific situation (e.g. body weight for allometric scaling).

Subsequently, using the tool APROBA, these distributions are combined in a probabilistic way and, as an outcome of this assessment, coverage and degree of uncertainty of the RfD are calculated, as well as a “probabilistic” RfD – that is, an  $HD_M^1$  corresponding to a predefined coverage (here: 95%).

The results of this probabilistic analysis are then compared with the expected exposure of the human target population, showing that the first tier does not provide a sufficiently conclusive answer to the question of whether current exposure to DON can be regarded as acceptable with respect to potential health effects in the human population.

One option at this point is to generate additional data aimed at reducing uncertainty. Another option, which is illustrated here, is to use a higher-tier method of dose–response analysis – that is, the BMD instead of the NOAEL approach. This higher-tier assessment can be performed with the same available database and is discussed in section A5.3.

Finally, an analogous assessment is presented for additional end-points in section A5.4.

## A5.2 Tier 1: Hazard characterization for general toxicity after repeated dosing using the NOAEL and default assessment factors

### **A5.2.1 Hazard characterization aspects**

As a starting point, hazard assessment and associated uncertainty analysis are performed for the effects related to general toxicity. The study selected for such effects was a 2-year dietary study in B6C3F1 mice (Iverson et al., 1995).

DON dose levels were 0, 0.12, 0.7 and 1.5 mg/kg bw per day. Group mean body weight as averaged over the whole study duration showed a clear dose-related reduction in treated groups compared with controls, in the absence of differences in feed consumption. Changes in organ weights could be explained by the reduced body weights in the treated groups. All other end-points considered showed little or no change. Therefore, body weight is the only end-point selected for further analysis. The details of the dose–response relationship regarding this effect are given in Table A5.10 (in Appendix I to this annex).

Reduced body weight can be considered as a general indicator of toxicity. Regarding the size of change in body weight considered to be adverse, a 5% change was used as the critical effect size or BMR by the authors of the original risk assessment (Pieters et al., 2001; Pieters, Bakker & Slob, 2004; Bokkers et al., 2009), and this is also followed in the present example.

#### *A5.2.1.1 Aspect I: Setting the PoD*

The data for decreased body weight from Iverson et al. (1995) are given in Table A5.10 in Appendix I to this case-study. Females of group B (the group receiving the lowest dose) showed a statistically significant ( $P < 0.01$ ) reduction in average body weight over lifetime versus controls. Therefore, a NOAEL cannot be identified from these data, and the lowest dose tested of 0.12 mg/kg bw per day or 120 µg/kg bw per day must be regarded as a LOAEL.<sup>24</sup>

A primary recommendation of the guidance in this situation is to perform a BMD analysis in all cases (section 4.2.3 of main text). Nevertheless, as NOAELs are still frequently used, the first tier in this example will be based on the NOAEL approach. This has the added benefit of demonstrating the reduction in uncertainty achieved by progressing from the NOAEL to the BMD approach.

#### *A5.2.1.2 Aspect II: Adjustment of the PoD (LOAEL à NOAEL)*

Currently, default factors for LOAEL to NOAEL extrapolation are commonly used. According to ECHA (2012),<sup>25</sup> defaults used by different international institutions (e.g. WHO/IPCS, USEPA, the European Centre for Ecotoxicology and Toxicology of Chemicals [ECETOC] or the European Chemicals Agency itself) vary, with 3 and 10 being the most

---

<sup>24</sup> The NOAEL approach contains ambiguity in the sense that some risk assessors would identify a NOAEL based on statistical significance only, whereas others would consider the observed difference in the mean responses to be more important than statistical significance. Whereas the former approach suffers from the limitation that the potential size of the effect is not taken into account, the latter is inappropriate for ignoring the statistical uncertainty in the data. Whereas this divergent practice points at some of the inherent limitations of the NOAEL approach, this issue will not be further discussed in the present document.

<sup>25</sup> See sections R.8.2 and R.8.4.3.1 of ECHA (2012) for a more extensive discussion of the problems associated with LOAEL to NOAEL extrapolation. Also see section 5.1.3 of the main text.

common values. However, as discussed in the main text, using default factors for extrapolating from LOAELs to NOAELs without considering the potential size of the effect at the LOAEL is hard to defend. In the case of DON, the available dose–response data do provide some basis for selecting a reasonable factor. In the present case, given that only a mild effect on body weight is observed at the LOAEL, and based on visual inspection of the overall dose–response data, it is estimated that a 3-fold lower dose would not have resulted in a (statistically and/or biologically) significant difference in body weight compared with controls had it been tested. Therefore, an assessment factor of 3 is used to extrapolate from the LOAEL to a NOAEL.

*A5.2.1.3 Aspect III: Interspecies extrapolation*

A default factor of 10 will be applied to the extrapolated NOAEL to extrapolate from mice to humans.

*A5.2.1.4 Aspect IV: Intraspecies extrapolation*

A default factor of 10 will be applied to the extrapolated NOAEL to account for interindividual variability.

*A5.2.1.5 Assessment factors and derivation of the reference dose*

Table A5.1 summarizes the assessment factors chosen in this example for the various hazard characterization aspects involved as well as the resulting RfD.

**Table A5.1: Summary of approaches chosen for each hazard characterization aspect**

<b><i>Hazard characterization aspect</i></b>	<b><i>AF</i></b>	<b><i>Calculation of RfD (µg/kg bw per day)</i></b>
I. PoD: LOAEL	–	120
II. LOAEL to NOAEL	3	40
III. Interspecies	10	4
IV. Intraspecies	10	<b>0.4<sup>a</sup></b>

AF: assessment factor; LOAEL: lowest-observed-adverse-effect level; NOAEL: no-observed-adverse-effect level; PoD: point of departure; RfD: reference dose

<sup>a</sup> The final RfD for the human target population is printed in bold.

***A5.2.2 Evaluation of uncertainties***

The evaluation of the uncertainties is performed by applying the approximate probabilistic method (see section 3.3.1 of the main text), using the APROBA tool. The APROBA tool contains two blocks of input fields:

1. inputs related to study, end-point and protection goal; and
2. inputs related to adjustment, variability and uncertainty.

This section illustrates what inputs need to be entered in both blocks.

## A5.2.2.1 Inputs related to study, end-point and protection goal

The first block is shown in Fig. A5.1. The third column provides various default values (if applicable), but the user can change any of those in the second column (all input cells in APROBA are yellow).

INPUTS RELATED TO STUDY, END-POINT AND PROTECTION GOALS		
DESCRIPTION	INPUTS	COMMON VALUE(S)
End-point	Reduced average lifetime bw	Case-specific
Data type	Continuous	Case-specific
Data route	Oral	Case-specific
Study type	Chronic	Case-specific
Test species	Mouse	Case-specific
Body weight test species (kg)	0.040	0.02
Human median body weight (kg)	50.0	60
Target BMR (= $M$ , user input for BMDLs only)	5%	5%
Population incidence goal (= $I$ )	1%	5%, 1%, 0.1%, 0.01%
Probabilistic coverage goal	95%	95%
PoD type	LOAEL	Case-specific
PoD value	120	Case-specific
BMDU (User input for BMDL PoDs)		Case-specific
PoD units	$\mu\text{g}/\text{kg}$ body weight per day	$\text{mg}/\text{kg}$ body weight per day
Deterministic overall AF	300	Case-specific
Deterministic RfD	0.40	Calculated
Exposure estimate (optional)	0.44	User supplied

Fig. A5.1: APROBA input section related to study, end-point and protection goals. The cells in yellow indicate user inputs.

Up to the row “Test species”, entries are self-evident. In the next two rows, an average body weight of mice of 40 g (estimated from the body weight curves in the study by Iverson et al., 1995) and an average body weight of the human target population (young individuals in growing phase, i.e. aged 0–19 years) of 50 kg are used as inputs. The target human dose in this example is defined as the (true) dose where 1% (row “Probabilistic incidence goal”) of the exposed target population would be subject to a 5% (row “Target BMR”) decrease in mean body weight compared with the mean body weight of the non-exposed population (i.e. here the target  $\text{HD}_M^I$  is defined as  $\text{HD}_{05}^{01}$ ).

In addition, a value of 95% coverage (row “Probabilistic coverage goal”) is specified for the subsequent calculation of a probabilistic RfD (see below).

In the next two rows the type and value of the PoD is entered, in this case a LOAEL of 120  $\mu\text{g}/\text{kg}$  bw per day. With an overall assessment factor of 300, the deterministic RfD is calculated in the second-last row of Fig. A5.1. Finally, the last row offers the possibility to enter a deterministic exposure estimate for the respective target population (in this case: 0.44  $\mu\text{g}/\text{kg}$  body weight per day).

A5.2.2.2 Inputs related to adjustment, variability and uncertainty

Fig. A5.2 shows the block of input fields provided in APROBA for entering the assumed uncertainty distributions. Unless indicated otherwise, generic default uncertainty distributions are used for the aspects involved, as proposed in Table 4.6 of the main text. Note that the large number of decimal places in some of the cells result from underlying calculations and may be ignored.

(a) Point of departure (PoD)

The first two rows (below the headers) in Fig. A5.2 are used for BMD uncertainty, which does not apply in this case. Therefore, both cells are given the value of 120 µg/kg bw per day (the value of the LOAEL in this case). The extrapolation of the LOAEL to the NOAEL is discussed separately below. In the next two rows, the uncertainty in the hypothetical NOAEL (after LOAEL to NOAEL extrapolation) is expressed, i.e. the uncertainty as an estimate of the BMDL, together with the assumed uncertainty in the hypothetical BMD, which is represented by the lognormal distribution for continuous data (with an LCL of 0.0709 and a UCL of 1.57), as proposed in Table 4.6 of the main text.

INPUTS RELATED TO ADJUSTMENT, VARIABILITY AND UNCERTAINTY			
HAZARD CHARACTERIZATION		INPUTS	PROVISIONAL VALUE(S)
<b>PoD</b>	LCL	120	Calculated from inputs
(Modelled BMD uncertainty)	UCL	120	Calculated from inputs
<b>NOAEL to BMD</b>	LCL	0.0709	0.0709
(NOAEL or LOAEL only)	UCL	1.57	1.57
<b>Interspecies scaling</b>	LCL	6.39	6.39
(Allometric for oral)	UCL	11.3	11.3
<b>Interspecies TK/TD</b>	LCL	0.333	0.33
(Remaining TK & TD)	UCL	3.00	3.00
<b>Duration extrapolation</b>	LCL	1.00	1.00
	UCL	1.00	1.00
<b>Intraspecies</b>	LCL	2.24	2.24
	UCL	41.9	41.9
<b>Other aspect #1</b>	LCL	0.67	1.00
LOAEL to NOAEL	UCL	6.00	1.00
<b>Other aspect #2</b>	LCL	1.00	1.00
(Description here)	UCL	1.00	1.00
<b>Other aspect #3</b>	LCL	1.00	1.00
(Description here)	UCL	1.00	1.00

**Fig. A5.2: APROBA input section related to adjustment, variability and uncertainty.** The cells in yellow indicate the user inputs. The last column provides suggestions for these inputs based on the distributions in the main text (section 4). LCL and UCL are lower (5%) and upper (95%) confidence limits.

(b) Interspecies extrapolation

In the deterministic calculation in Table A5.1, a default factor of 10 was applied, without allometric scaling.<sup>26</sup> Here, the uncertainty in interspecies extrapolation is based on the generic uncertainty distribution for allometric scaling (given the specific animal body weight)

<sup>26</sup> Of course, allometric scaling can also be applied in a deterministic approach. In this case, where the test animals were mice, this would – together with remaining TK/TD differences – result in an adjustment factor higher than 10 (note that the allometric scaling factor increases with larger interspecies difference in body weight).

and the generic uncertainty distribution for remaining TK/TD differences (see Table 4.6 of the main text).

Uncertainty in allometric scaling is represented by the default normal distribution for the power with mean = 0.7, SD = 0.024. As mentioned above, a body weight of 40 g for the test animals and a body weight of 50 kg for the target human population (young people) are assumed. This results in a lognormal scaling factor with an LCL of approximately 6.4 and a UCL of 11.3 (rows “Interspecies scaling/(Allometric for oral)”; these values are calculated automatically by the APROBA tool once the user specifies the body weights in the input section).

The TK/TD differences are represented by the lognormal distribution with GM = 1 and P95/P50 = 3 – that is, with an LCL of 1/3 and a UCL of 3 (rows “Intraspecies scaling/(Remaining TK & TD)”; see section 4.4.2 of main text).

(c) Intraspecies extrapolation

The intraspecies uncertainty relates to the factor between the P01 (1st percentile) of the population and the P50 (the median or typical individual) – that is, it relates to  $I = 1\%$  as given in Fig. A5.1. This uncertainty is represented by a lognormal uncertainty distribution for that incidence with LCL = 2.2 and UCL = 42, using the results in section 4.5.2 of the main text (row “Intraspecies” and next row).

(d) LOAEL to NOAEL

Given that there can be no default value for the uncertainty in the LOAEL to NOAEL extrapolation, the APROBA tool does not provide a generic default distribution for this aspect, and an assessment needs to be made on a case-by-case basis. In this case, the LOAEL of 120  $\mu\text{g}/\text{kg}$  bw per day for body weight reduction was adjusted by a factor of 3, for the reasons given in section A5.2.1.2 above. However, this factor was considered to be a conservative estimate of an uncertainty factor. The latter uncertainty might be represented by a lognormal distribution with GM = 2 and P95/P50 = 3, and the resulting confidence interval is (0.67, 6.0). Thus, it is assumed that the NOAEL is unlikely to be more than 6 times lower than this LOAEL (row “Other aspect #1” and “LOAEL to NOAEL”; note that according to the assumed confidence interval, the NOAEL could theoretically also have been slightly higher than this LOAEL with small probability; this reflects the possibility that the LOAEL dose group in females was an outlier due to some experimental factor other than the dose).

**A5.2.2.3 Results**

The results of the probabilistic analysis are summarized in Fig. A5.3, by showing the corresponding output table of the APROBA tool.

According to this analysis, the deterministic RfD of 0.4  $\mu\text{g}/\text{kg}$  bw per day has a coverage of only about 85%. In other words, there is a 15% chance that the “true”  $\text{HD}_{05}^{01}$  will be lower than the deterministic RfD.

The probabilistic RfD is calculated to be about 0.15  $\mu\text{g}/\text{kg}$  bw per day. In other words, 0.15  $\mu\text{g}/\text{kg}$  bw per day is the human dose at which 1% of the population (growing individuals) would be subject to a  $\geq 5\%$  reduction in body weight with 95% coverage (or, equivalently, the human dose at which 99% of the population would be subject to less than a 5% reduction in body weight is estimated to be 0.15  $\mu\text{g}/\text{kg}$  bw per day, with 95% coverage).

The degree of uncertainty – that is, the ratio between the upper and the lower confidence limits associated with this  $HD_{05}^{01}$  – is about a factor of 200. In other words, the true  $HD_{05}^{01}$  could be considerably higher, up to around 200-fold (for the confidence level used), than the probabilistic RfD.

NON-PROBABILISTIC ANALYSIS OUTPUTS <sup>l,k</sup>			
Target Human Dose ( $HD_M^l$ )	LCL	0.00899	µg/kg bw per day
	UCL	532	µg/kg bw per day
Fold Range of Uncertainty		59187	
Estimated "Coverage" of Non-Prob. LCL of $HD_M^l$ *			100.0%
*Based on approximate probabilistic analysis, below.			
APPROXIMATE PROBABILISTIC ANALYSIS OUTPUTS			
Standard Confidence Interval			
Target Human Dose ( $HD_M^l$ )	LCL (P05)	0.154	µg/kg bw per day
	UCL (P95)	31.0	µg/kg bw per day
Degree of Uncertainty (Fold Range)			201
Estimated "Coverage" of Deterministic RfD			85%
Probabilistic RfD	= Approximate probabilistic $HD_M^l$ at specified % confidence		
0.15	= Estimate of dose (µg/kg bw per day) at which, with		
	95%	confidence	
	1%	of the population will have	Example calculation
	of magnitude	≥	5%

Fig. A5.3: Approximate probabilistic results as provided by APROBA related to a deterministic hazard characterization using the NOAEL approach, with body weight as the critical end-point

A5.2.2.4 Is a higher-tier assessment needed?

(a) General considerations

The decision as to whether the results from the first-tier assessment are sufficient or whether a higher tier would be needed cannot be separated from risk management considerations. Which coverage or degree of uncertainty is sufficient in the context of a given hazard characterization, which margin of exposure might be adequate or which percentile of the population should be protected from which degree of effect are all decisions that should be made in collaboration between assessors and regulators. The discussion below serves only as an illustration of the argumentation behind answering the question, “Do we know enough, or is a higher-tier assessment needed?”

As mentioned in the previous section, the RfD coverage for the body weight effects calculated at tier 1 is low, implying a relatively large risk that the RfD does not fulfil the protection goals. Furthermore, the estimated exposure level of the human target population in this case is 0.44 µg/kg bw per day (see RIVM publications given in section A5.1.1), and the non-probabilistic RfD is very close to this level.

Further, the probabilistic RfD with 95% coverage (0.15 µg/kg bw per day) is almost 3-fold lower than human exposure (0.44 µg/kg bw per day), suggesting that this exposure might not comply with the protection goals as defined by *M* and *I*. However, the high degree of uncertainty indicates that the target human dose could be much higher (up to around 200-fold) than the probabilistic RfD.

Taking all of these considerations together, it may be concluded that this first-tier assessment was unable to establish with adequate confidence that the protection goals are met, whereas the uncertainty in the  $HD_M^I$  was very large. Therefore, it is decided to perform a higher-tier assessment.

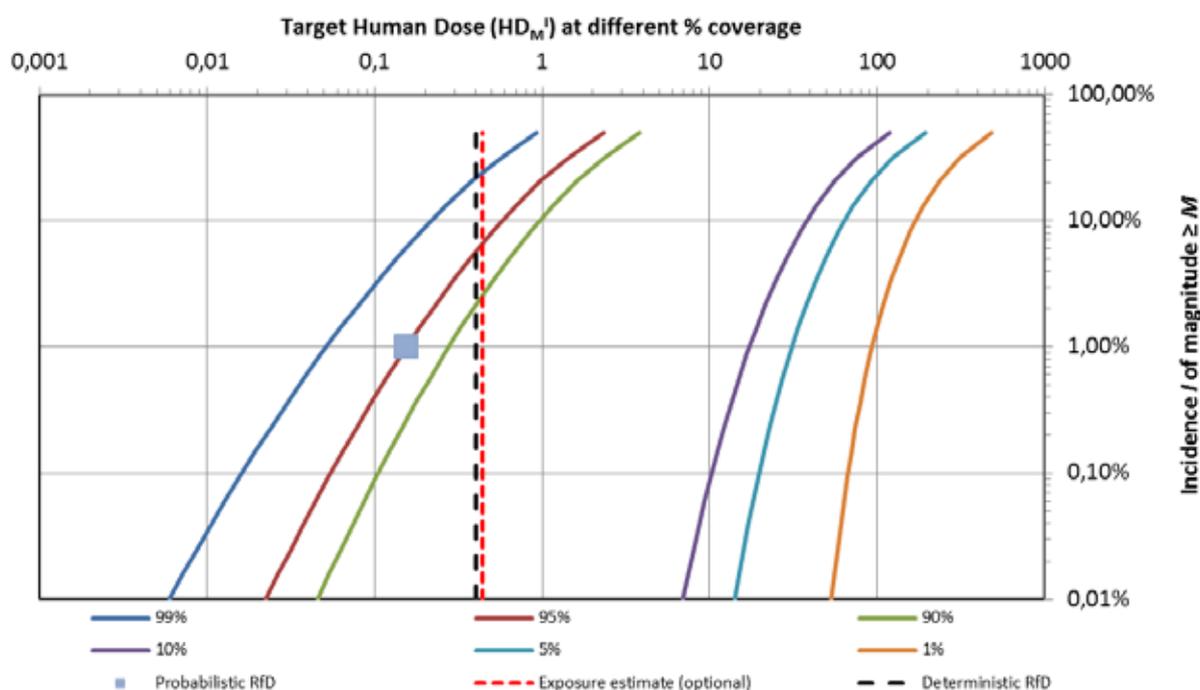
(b) Additional uncertainties

In the evaluation of uncertainties above, only those aspects that were also considered in the deterministic assessment were taken into account. In principle, additional uncertainties should be evaluated as well. However, in this particular case, the decision was made to perform a higher-tier assessment, and a further evaluation of additional uncertainties would not change that and will be omitted here. The evaluation of additional uncertainties will be illustrated below (see sections A5.3.4 and A5.4.2.6).

(c) Additional output of APROBA

Before addressing the question as to which type of higher-tier assessment would be desirable, we first discuss some additional output from APROBA, as part of this output is helpful in deciding how to improve the assessment in a higher tier.

The APROBA tool offers a visual means of evaluating the outcome of the approximate probabilistic analysis. Fig. A5.4 illustrates this graphical output, which shows the confidence interval of the estimated  $HD_M^I$  as a function of the incidence  $I$  associated with that  $HD_M^I$ .



**Fig. A5.4: APROBA graphical output for body weight effects using the NOAEL approach.** The x-axis represents human dose on a log scale (see numbers at the top), the y-axis the incidence for effect  $M$ . The three curves at the left represent lower-bound estimates of  $HD_M^I$ , the three curves on the right the upper-bound estimates. The associated coverages are indicated at the bottom. The vertical dashed line indicates the deterministic RfD, the dotted line the estimated human exposure.

The square dot in Fig. A5.4 represents the lower-bound estimate of the  $HD_{05}^{01}$  – that is,  $0,15 \mu\text{g}/\text{kg bw}$  per day (coverage = 95%). Moving along the curve (carrying the square dot) to the right-hand side increases the lower-bound estimate of the  $HD_M^I$  (with coverage still at

95%) at the expense of an increase in incidence (e.g. going up to an  $HD_M^I$  of 0.6  $\mu\text{g}/\text{kg}$  bw per day corresponds to an estimated increase in incidence from 1% to 10%). Keeping the lower-bound estimate of  $HD_M^I$  constant (i.e. moving vertically in Fig. A5.4) changes coverage and incidence at the same time. For example, the human dose of 0.15  $\mu\text{g}/\text{kg}$  bw per day (square dot) could also represent a lower-bound estimate of the human dose with about 6% incidence with 99% coverage. Finally, moving horizontally in Fig. A5.4 keeps incidence constant and demonstrates the effect of changing the lower-bound estimate of the  $HD_M^I$  on the associated coverage.

From the above representation, it is evident that for the target incidence of 1%, coverage of the deterministic RfD is comparatively low (from Fig. A5.3 above, we already know it to be around 85%). The same dose would correspond to a calculated incidence of about 2% when coverage is set to 90% (hypothetical intersection of dashed line and graph for 90% coverage in Fig. A5.4). For a coverage of 99%, the associated incidence would be  $> 20\%$  (intersection of dashed line and graph for 99% coverage in Fig. A5.4).

In Fig. A5.4, the vertical dotted line represents the estimated exposure level of the target population (0.44  $\mu\text{g}/\text{kg}$  bw per day). As stated previously, the dashed line next to it marks the deterministic RfD of 0.4  $\mu\text{g}/\text{kg}$  bw per day. Thus, it is immediately evident that estimated exposure is higher than both calculated RfDs (deterministic and probabilistic). Note, however, that the extent to which exposure exceeds the deterministic RfD is negligible in light of the uncertainties reflected by the confidence interval curves.

Finally, when interpreting Fig. A5.4, it should be borne in mind that the confidence intervals in the graph do not include potential uncertainties that were not evaluated.

Another feature of the APROBA software is an output table, which indicates the contributions of the uncertainties in each aspect of the hazard characterization to the overall uncertainty. Obviously, these results are again approximations, as they are based on an approximate probabilistic assessment. The output provided by APROBA is illustrated in Fig. A5.5.

The table shown in Fig. A5.5 is very helpful for identifying those aspects of the specific hazard characterization on which attention should be focused in a higher-tier assessment in order to reduce the overall uncertainty. Of course, in selecting the aspect most suitable for refinement, other considerations will normally be relevant as well (e.g. feasibility, costs).

In this example, it is evident that the two aspects associated with the PoD – that is, NOAEL to BMD extrapolation (34%) and LOAEL to NOAEL extrapolation (“Other aspect #1”, 17%) – are together responsible for the greatest part of the uncertainty in the calculated  $HD_M^I$ .

#### *A5.2.2.5 How to proceed?*

As explained above (section A5.2.2.4), it may be concluded that this first-tier assessment did not result in a positive answer to the question “Do we know enough?” and that a higher-tier assessment would be desirable. As the current assessment was based on the NOAEL approach, the higher-tier assessment to be considered as the first option is one that reanalyses the data with a better statistical method in order to arrive at a less uncertain PoD: the BMD approach. This may be expected to result in a smaller degree of uncertainty associated with the derived RfD. This approach is clearly much cheaper and quicker than other options, such as generating additional data with respect to, for example, chemical-specific interspecies TK/TD extrapolation or intraspecies variability.

INTERMEDIATE CALCULATIONS FOR UNCERTAINTY ANALYSES				% contribution
ASPECT			$[\log(P95/P50)]^2$	to overall uncertainty
PoD	P50	120		--
	P95/P50	1.00	0	
NOAEL to BMD	P50	0.333		34%
	P95/P50	4.70	0.452	
Interspecies scaling	P50	8.49		1%
	P95/P50	1.33	0.0153	
Interspecies TK/TD	P50	1.00		17%
	P95/P50	3.00	0.228	
Duration extrapolation	P50	1.00		--
	P95/P50	1.00	0	
Intraspecies	P50	9.69		30%
	P95/P50	4.32	0.404	
Other aspect #1 LOAEL to NOAEL	P50	2.00		17%
	P95/P50	3.00	0.228	
Other aspect #2 (Description here)	P50	1.00		--
	P95/P50	1.00	0	
Other aspect #3 (Description here)	P50	1.00		--
	P95/P50	1.00	0	
Target Human Dose ( $HD_M$ )		Non-Prob.	Approx. Prob.	Greatest contributor
	P50	2.19	2.19	to overall uncertainty
	UCL/P50	243	14.2	NOAEL to BMD

Fig. A5.5: APROBA output showing the contribution to overall uncertainty of the individual aspects of hazard characterization for body weight using the NOAEL approach and default assessment factor

### A5.3 Tier 2: Hazard characterization for general toxicity after repeated dosing based on BMDL

In this section, the same data serve as the basis of the hazard characterization, but they are analysed by the BMD approach rather than the NOAEL approach. The other aspects are evaluated in the same way as in the previous section.

#### A5.3.1 Dose–response analysis for general toxicity (BMD approach)

The BMD analyses were performed using the software PROAST (<http://www.proast.nl>). Only the main results are reported here. For further details, see Appendix II of this case-study.

#### A5.3.2 BMDL for general toxicity

At the tier 1 level of this assessment (see sections A5.2.1 and A5.2.2), all treatment groups were tested against the controls, and it was found that the effects observed at the lowest dose in female mice (but not in male mice) differed significantly from those in the controls. However, in toxicity studies, individual dose groups can be subject to perturbations caused by experimental factors other than the dose or simply by random sampling error.

In the BMD approach, the hypothesis that males and females show the same dose–response relationship can be tested by considering the complete dose–response data set (see Table A5.10 in Appendix I). The BMD analysis resulted in significant differences between the two sexes in background body weights, as well as in the within-group variances. Males and females were, however, not found to differ significantly in sensitivity to DON for this endpoint (see Appendix II for a summary of the results). Both the exponential and the Hill models resulted in similar fits and similar confidence intervals for the BMD<sub>05</sub>.

The range from the lowest BMDL to the highest BMDU for the BMD at BMR = 0.05 was from 0.17 to 0.34 mg/kg bw per day. The BMDL for body weight is therefore set at 0.17 mg/kg bw per day. The width of the confidence interval was a factor of 2.

Using the same default assessment factors for interspecies and intraspecies extrapolation as in section A5.2.1 (10 each, 100 overall), a deterministic RfD of 1.7 µg/kg bw per day is obtained.

### **A5.3.3 Evaluation of uncertainties**

The uncertainties of the aspects involved are evaluated using the Microsoft Excel spreadsheet APROBA. The uncertainty distributions used for these aspects are the same as for the assessment based on the NOAEL in section A5.2, except that uncertainties associated with the NOAEL or LOAEL vanish and the uncertainties in the BMD derive from the specific data set considered rather than from a generic assumption.

The resulting APROBA input table is given in Fig. A5.6. The numerical output is presented in Fig. A5.7.

As expected from the results in Fig. A5.7, reducing the uncertainty in the PoD markedly reduced the overall uncertainty in the RfD. Table A5.2 summarizes what has been gained by moving from tier 1 to tier 2.

Replacing the LOAEL/NOAEL with the BMDL allowed the deterministic RfD to be increased more than 4-fold. At the same time, the degree of uncertainty in the RfD decreased by almost 5-fold. The probabilistic RfD (with 95% coverage) was raised by almost 3-fold and is now in the same range as the expected exposure of the target population. These results are further illustrated in APROBA's graphical output, shown in Fig. A5.8.

INPUTS RELATED TO STUDY, END-POINT AND PROTECTION GOALS			
DESCRIPTION		INPUTS	COMMON VALUE(S)
End-point		Reduced average lifetime bw	Case-specific
Data type		Continuous	Case-specific
Data route		Oral	Case-specific
Study type		Chronic	Case-specific
Test species		Mouse	Case-specific
Body weight test species (kg)		0.040	0.020
Human median body weight (kg)		50.0	60.0
Target BMR (= <i>M</i> , user input for BMDLs only)		5%	5%
Population incidence goal (= <i>I</i> )		1%	5%, 1%, 0.1%, 0.01%
Probabilistic coverage goal		95%	95%
PoD type		BMDL	Case-specific
PoD value		170	Case-specific
BMDU (User input for BMDL PoDs)		340	Case-specific
PoD units		µg/kg bw per day	mg/kg body weight per day
Deterministic overall AF		100	Case-specific
Deterministic RfD		1.70	Calculated
Exposure estimate (optional)		0.44	User supplied
INPUTS RELATED TO ADJUSTMENT, VARIABILITY AND UNCERTAINTY			
HAZARD CHARACTERIZATION		INPUTS	PROVISIONAL VALUE(S)
PoD	LCL	170	Calculated from inputs
(Modelled BMD uncertainty)	UCL	340	Calculated from inputs
NOAEL to BMD	LCL	1.00	1.00
(NOAEL or LOAEL only)	UCL	1.00	1.00
Interspecies scaling	LCL	6.39	6.39
(Allometric for oral)	UCL	11.3	11.3
Interspecies TK/TD	LCL	0.333	0.33
(Remaining TK & TD)	UCL	3.00	3.00
Duration extrapolation	LCL	1.00	1.00
	UCL	1.00	1.00
Intraspecies	LCL	2.24	2.24
	UCL	41.9	41.9
Other aspect #1	LCL	1.00	1.00
(Description here)	UCL	1.00	1.00
Other aspect #2	LCL	1.00	1.00
(Description here)	UCL	1.00	1.00
Other aspect #3	LCL	1.00	1.00
(Description here)	UCL	1.00	1.00

Fig. A5.6: APROBA input table for uncertainty characterization regarding body weight effects and using the BMD approach

NON-PROBABILISTIC ANALYSIS OUTPUTS <sup>j,k</sup>			
Target Human Dose (HD <sub>M</sub> <sup>l</sup> )	LCL	0.120	µg/kg body weight per day
	UCL	71.3	µg/kg body weight per day
Fold Range of Uncertainty		595	
Estimated "Coverage" of Non-Prob. LCL of HD <sub>M</sub> <sup>l*</sup>			100%
*Based on approximate probabilistic analysis, below.			
APPROXIMATE PROBABILISTIC ANALYSIS OUTPUTS			
Standard Confidence Interval			
Target Human Dose (HD <sub>M</sub> <sup>l</sup> )	LCL (P05)	0.444	µg/kg body weight per day
	UCL (P95)	19.2	µg/kg body weight per day
Degree of Uncertainty (Fold Range)			43
Estimated "Coverage" of Deterministic RfD			68%
Probabilistic RfD	= Approximate probabilistic HD <sub>M</sub> <sup>l</sup> at specified % confidence		
0.44	= Estimate of dose (µg/kg body weight per day) at which, with		
	95%	confidence	
	1%	of the population will have	Reduced average lifetime bw
	of magnitude	≥	5%

Fig. A5.7: APROBA numerical output for uncertainty characterization regarding body weight effects and using the BMD approach

Table A5.2: Summary of uncertainty evaluation of BMDL-based assessments

	Deterministic RfD (µg/kg bw per day)	Coverage (%)	Degree of uncertainty	Probabilistic RfD <sup>a</sup> (µg/kg bw per day)	Exposure of target population (P95) (µg/kg bw per day)
Tier 1	0.4	85	201	0.15	
Tier 2	1.7	68	43	0.44	0.44

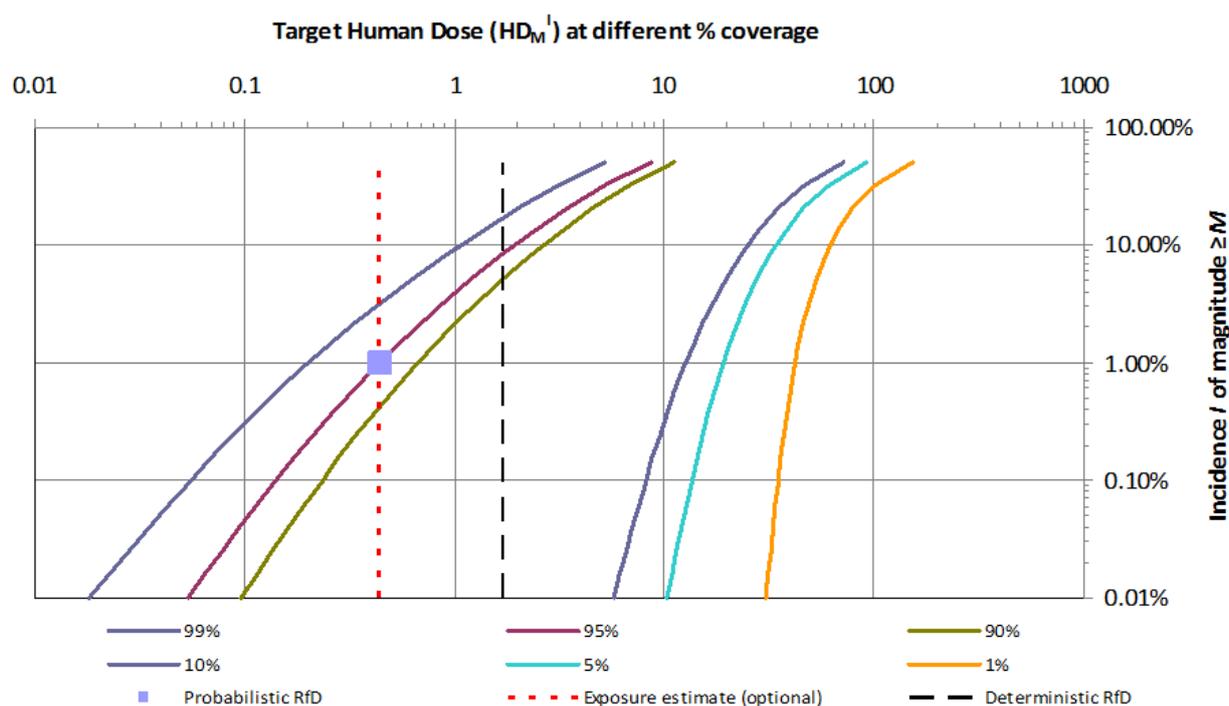
bw: body weight; P95: 95th percentile; RfD: reference dose

<sup>a</sup> For incidence *I* = 1% and coverage = 95%.

#### A5.3.4 Is a higher-tier assessment needed?

For body weight reduction, there is a considerable margin between the deterministic RfD and human exposure, but the coverage of the deterministic RfD is fairly low (around 68%). For a coverage of 95%, the probabilistic RfD is equal to the human exposure level. Therefore, if the associated protection goal (*M* = 5% body weight reduction, *I* = 1%) is considered acceptable, this outcome might be considered sufficient to conclude that an unacceptable health effect on body weight is not expected, given the estimated exposure.

However, any additional uncertainties that have not been quantitatively evaluated in APROBA need to be considered, particularly given the lack of any margin between exposure and the probabilistic RfD. For instance, the PoD was derived from a single chronic study, and it remains unknown how large the potential interstudy variation might be (see section 3.4.4 of the main document).



**Fig. A5.8: APROBA graphical output for body weight effects using the BMD approach.** The x-axis represents human dose on a log scale (see numbers at the top), the y-axis the incidence for effect  $M$ . The three curves at the left represent lower-bound estimates of  $HD_M^I$ , the three curves on the right the upper-bound estimates. The associated coverages are indicated at the bottom. The vertical dashed line indicates the deterministic RfD, the dotted line the estimated human exposure.

Further, it might be that some of the uncertainties that were quantitatively evaluated used generic uncertainty distributions that were based on limited or not completely representative data. Such uncertainties (“secondary” uncertainties; see section 3.4.1 of the main text) were generically evaluated in Table 4.8 of the main text. Here, it was indicated that for some aspects, there might be reason for a further case-specific evaluation. This does not seem to apply to the current case, because, for example, effects on body weight were well covered by the data sets used to inform the generic distributions for interspecies uncertainty. Therefore, it might be decided that secondary uncertainties do not need further attention in the present case. If so, the risk manager might conclude that no higher-tier assessment would be needed based on this specific hazard characterization (and body weight effects as the critical endpoint).

If, however, the risk manager decided that a more refined assessment was needed, it is evident that, based on Fig. A5.9, generating data for the development of chemical-specific adjustment factors for intraspecies and/or interspecies variability would appear to be the most promising step forward in this case, as they are responsible for more than 90% of the overall uncertainty.

INTERMEDIATE CALCULATIONS FOR UNCERTAINTY ANALYSES				% contribution
ASPECT			$[\log(P95/P50)]^2$	to overall uncertainty
PoD	P50	240		3%
	P95/P50	1.41	0.0227	
NOAEL to BMD	P50	1.00		--
	P95/P50	1.00	0	
Interspecies scaling	P50	8.49		2%
	P95/P50	1.33	0.0153	
Interspecies TK/TD	P50	1.00		34%
	P95/P50	3.00	0.228	
Duration extrapolation	P50	1.00		--
	P95/P50	1.00	0	
Intraspecies	P50	9.69		60%
	P95/P50	4.32	0.404	
Other aspect #1 (Description here)	P50	1.00		--
	P95/P50	1.00	0	
Other aspect #2 (Description here)	P50	1.00		--
	P95/P50	1.00	0	
Other aspect #3 (Description here)	P50	1.00		--
	P95/P50	1.00	0	
(HD <sub>M</sub> ) <sup>1</sup>		Non-Prob.	Approx. Prob.	Greatest contributor
	P50	2.92	2.92	to overall uncertainty
	UCL/P50	24.4	6.58	Intraspecies

Fig. A5.9: APROBA tabular output showing the contribution to overall uncertainty of the individual aspects of hazard characterization using the BMD approach

#### A5.4 Including further end-points

Aside from the effect on body weight addressed in the previous sections, there are other important effects of DON found in animal studies, in particular those related to reproductive toxicity. An overview of the critical studies and possible candidate end-points for derivation of an RfD is given in Table A5.3. These end-points are discussed in this section.

Table A5.3: Summary of the studies and end-points potentially used for RfD derivation for DON

Effect	Candidate end-points	Study
General toxicity	Mean body weight over lifetime	Diet, 2 years in mice; Iverson et al. (1995)
Development	Fetal weight	Gavage, gestation days 6–19 in rats; Collins et al. (2006)
	Resorptions, fetal anomalies	Gavage, gestation days 8–11 in mice; Khera et al. (1982)
Male fertility	Epididymal weight, seminal vesicle weight, testicular sperm count, germ cell degeneration, failure of sperm release	Gavage, 28 days in rats; Sprando et al. (2005)

**A5.4.1 Tier 1 assessment of end-points of reproductive toxicity using the NOAEL approach and default assessment factors**

**A5.4.1.1 Candidate studies**

(a) Prenatal development

The two prenatal developmental toxicity studies chosen here are Khera et al. (1982, gavage, mice) and Collins et al. (2006, gavage, rats).

In the prenatal developmental toxicity study by Khera et al. (1982), doses of 0, 0.5, 1, 2.5, 5, 10 and 15 mg/kg bw per day were administered to female mice via gavage (in fact, the study had two parts that were carried out sequentially; the first segment included dose levels of 5 mg/kg bw per day and above, and the second segment contained the lower doses). Aside from not having been performed to good laboratory practice (GLP) standards, there are strong deviations from OECD test guideline 414 (as issued in 1981). The number of treated dams (varying from 15 to 19) per group was lower than required by the guideline, and these animals were treated only on gestation days 8–11 instead of days 6–15 as recommended. Thus, the whole period of organogenesis has not been covered in this study. Moreover, data on anomalies and resorptions are reported insufficiently according to today's standards (e.g. the number of litters affected by at least one anomaly or the percentage of resorptions compared with total implants is given for each dose group, but it is not clear how these parameters were distributed over individual litters). Also, no information on maternal well-being (other than body weight development) is given.

For this assessment, the effects to be considered are pronounced dose-related increases in both the frequency of resorptions and the number of anomalous fetuses. Details of the corresponding dose–response relationships are given in Tables A5.11 and A5.12 in Appendix I. Both effects are considered directly relevant to humans; for the determination of NOAELs/LOAELs, any statistically significantly increased incidence compared with controls is considered adverse.

The following PoDs were identified:

- For incidence of resorptions, the dose level of 2.5 mg/kg bw per day differed significantly from the controls, and the NOAEL was 1 mg/kg bw per day.
- For incidence of anomalous fetuses, the dose level of 1 mg/kg bw per day differed significantly from the controls, and the NOAEL was 0.5 mg/kg bw per day.

Therefore, the lowest NOAEL for developmental effects (i.e. 0.5 mg/kg bw per day) is associated with fetal anomalies in mice.

In the prenatal developmental toxicity study by Collins et al. (2006), female rats received a dose of 0, 0.5, 1, 2.5 or 5 mg/kg bw per day via gavage. The quality of this study generally appears to be high, its design was in line with OECD test guideline requirements and it was most likely performed to GLP standards. Fetal weight appeared to be the most sensitive end-point.

For this assessment, a 5% decrease in fetal weight is considered to be adverse. With significant changes at the two highest dose levels (see Table A5.13 in Appendix I), the NOAEL for this effect was 1 mg/kg bw per day in both males and females.

(b) Male fertility

For male fertility, the only study available is Sprando et al. (2005), a 28-day study in which doses of 0, 0.5, 1, 2.5 and 5 mg/kg bw per day were administered to male rats (for details, see Table A5.14 in Appendix I).

Sprando et al. (2005) followed an individual test protocol, and the study does not relate to an existing OECD test guideline (but references to a detailed description of the data evaluation procedure are provided). Aside from body and organ weight measurements, the spectrum of observed effect parameters is limited to end-points of development, morphology and functionality of the male reproductive system. The study has most likely been performed to GLP standards, as the authors were affiliated with the United States Food and Drug Administration.

The following end-points, all related to male fertility, showed dose-related changes: epididymal weight, seminal vesicle weight, testicular sperm count, mild germ cell degeneration and failure of sperm release.

The first three of these end-points relate to continuous data, and in this assessment changes larger than 10% are considered adverse. The end-points germ cell degeneration and failure of sperm release were observed as quantal data, and any (statistically) significant increase in these findings compared with the controls is considered adverse.

*A5.4.1.2 Assessment factors*

(a) Aspect I: PoD

As NOAELs are available for all end-points considered here, no further adjustment is needed for this aspect.

(b) Aspect II: Interspecies extrapolation

A default factor of 10 for interspecies extrapolation is used and is directly applied to the PoDs.

(c) Aspect III: Intraspecies extrapolation

A default factor of 10 for intraspecies extrapolation is used and is directly applied to the PoDs.

(d) Aspect IV: Low study quality

As mentioned above, the study in mice by Khera et al. (1982) had serious flaws. In particular, DON was not administered over the whole period of organogenesis, and the reporting lacks sufficient detail for a complete assessment of the results (e.g. findings are not given on a per litter basis). The NOAEL for malformations or resorptions could have been lower had the whole period of organogenesis been covered or had per litter results been evaluated. Given the seriousness of the associated effects, an additional assessment factor of 5 is assigned to compensate for the lack of study quality in this example.

(e) Aspect V: Extrapolating short exposure duration in young males

An additional assessment factor of 10 will be applied to extrapolate from the subacute (28-day) study (Sprando et al., 2005) to a chronic (lifetime) exposure.

A5.4.1.3 Derivation of tier 1 reference doses

Table A5.4 shows the derivation of RfDs for reproductive toxicity obtained by applying the assessment factors to the NOAELs for the two effects. Although derived from different PoD values, developmental and fertility effects happen to result in the same tier 1 RfDs of 1 µg/kg bw per day (note that the most sensitive effect does not necessarily result in the lowest RfD).

**Table A5.4: Deriving RfDs based on the NOAEL approach**

	<i>Derivation of human RfDs (all values in µg/kg bw per day)</i>	
	<i>Development (anomalies in mice)</i>	<i>Fertility (of male rats)</i>
NO(A)EL	500	1000
Interspecies	50	100
Intraspecies	5	10
Data quality	<b>1<sup>a</sup></b>	–
Exposure duration	–	<b>1<sup>a</sup></b>

NO(A)EL: no-observed-(adverse-)effect level; RfD: reference dose

<sup>a</sup> The final RfD for the human target population is printed in bold.

A5.4.1.4 Evaluation of uncertainties

For the two reproductive end-points, the uncertainties are again evaluated using the tool APROBA. In all cases, the target human dose is defined as the (true) dose at which 1% of the population would be subject to the effect of magnitude  $M$  considered. So, the target  $HD_M^I$  has  $I = 1\%$  in all cases, whereas  $M$  depends on the end-point (as defined above).

The uncertainty distributions used for the aspects involved are as proposed in section 4 of the main text, unless indicated otherwise.

(a) Fetal anomalies in mice

PoD

Fetal anomalies are quantal data, for which the stochastic interpretation (see section 3.2.2 of the main text) will be used – that is, the incidences in the animal study are taken as reflecting the probability of anomalies in an individual dam/mother.<sup>27</sup> The NOAEL is considered to be an estimate of the  $BMDL_{10}$  (i.e. at  $BMR = 10\%$  extra risk); hence, the target human dose is defined as the dose at which 1% of the pregnant human mothers would have an extra risk of 10% or more of having a fetus with a developmental anomaly (i.e. 10% extra risk is the measure of severity  $M$ ).

The NOAEL was 0.5 mg/kg bw per day. The NOAEL uncertainty distribution (for stochastic quantal data) has an LCL (P05) of 0.14 and a UCL (P95) of 3.2 (see Table 4.6 of the main text).

<sup>27</sup> No dam-specific information was reported in this study, but it may be assumed that the mean frequency of anomalies at a given dose represents the probability of anomalies in the average dam.

*Interspecies extrapolation*

Uncertainty in allometric scaling is represented by the default distribution for the allometric power: normal with mean = 0.7, SD = 0.024. The allometric factor assumes a body weight of the test animal of 30 g (given in the study report) and a body weight of the target human of 60 kg, the latter being the assumed body weight of the median human female. This results in a lognormal distribution with an LCL of 7.2 and a UCL of 13 (the latter is calculated by APROBA).

For TK/TD differences, the default distribution has an LCL of 1/3 and a UCL of 3.

*Intraspecies extrapolation*

The default distribution has an LCL of 2.2 and a UCL of 42.

*Data quality*

An assessment factor of 5 was applied to correct for low data quality. This value was assumed to be conservative. Assuming that the factor might also have been close to 1 (i.e. same dose–response relationship if the study had been of good quality), the uncertainty range might be characterized by an LCL of 1 and a UCL of 5.

(b) Fertility

The Sprando et al. (2005) study resulted in a NOAEL of 1 mg/kg bw per day for two continuous end-points as well as for two quantal end-points. Therefore, the uncertainty evaluation will be done for epididymis weight and failure of sperm release, representing a continuous and a quantal end-point, respectively. Note that the results of the probabilistic analysis would be the same for the other critical end-points.

*PoD (epididymis weight)*

The uncertainty in the NOAEL (as an estimate of the BMDL<sub>05</sub>) is covered using the proposed uncertainty distribution in the main text (Table 4.6), the one for continuous chronic/subchronic data (LCL = 0.07, UCL = 1.6).

*PoD (failure of sperm release)*

Here, failure of sperm release is considered to be a stochastic quantal end-point, and the target human dose will be defined as the dose at which 1% of the male subpopulation would have a 10% probability of failure of sperm release, as the measure of severity *M*.

The NOAEL (as an estimate of the BMDL<sub>10</sub>) contains the regular uncertainty for the case of quantal data, which is represented by the lognormal distribution for stochastic quantal data (LCL = 0.14, UCL = 3.1).

*Interspecies extrapolation*

Uncertainty in allometric scaling is represented by the default distribution for the allometric power: normal with mean = 0.7, SD = 0.02. The allometric factor assumes a body weight of the test animal of 300 g and a body weight of the target human of 70 kg, the latter being the

assumed body weight of the median human male. This results in a lognormal scaling factor with an LCL of 4.1 and a UCL of 6.4 (calculated in APROBA).

For TK/TD differences, the default distribution has an LCL of 1/3 and a UCL of 3.

*Intraspecies extrapolation*

The default distribution has an LCL of 2.2 and a UCL of 42.

*Short exposure*

The critical study is a subacute study, whereas effects might occur after chronic exposure. This uncertainty is represented by the subacute to chronic uncertainty distribution, with an LCL of 5/8 and a UCL of 40 (see Table 4.6 of the main text).

(c) Results

The results of the APROBA calculations are given in Table A5.5. Differences in estimated exposure levels are explained by the target populations being different.

**Table A5.5: Summary of uncertainty evaluation of NOAEL-based assessments for further end-points**

	<i>Deterministic RfD (µg/kg bw per day)</i>	<i>Coverage (%)</i>	<i>Degree of uncertainty</i>	<i>Probabilistic RfD (µg/kg bw per day) (95% coverage)</i>	<i>Exposure in target population (P95)</i>
Development (fetal anomalies)	1	79	163	0.28	0.39 <sup>a</sup>
Fertility (epididymis weight)	1	90	579	0.50	0.84 <sup>b</sup>
Fertility (failure of sperm release)	1	82	579	0.25	0.84 <sup>b</sup>

bw: body weight; P95: 95th percentile; RfD: reference dose

<sup>a</sup> Females, aged 15–45 years.

<sup>b</sup> Males, aged 15–45 years.

*A5.4.1.5 Is a higher-tier assessment needed?*

(a) Fetal anomalies

The deterministic RfD associated with fetal anomalies is higher than human exposure, but the associated coverage is only around 79%. The estimate of the probabilistic RfD (with coverage 95%) is lower than the deterministic human exposure estimate (by a factor of 1.4). Furthermore, the associated target human dose relates to a 10% higher probability of anomalies (in 1% of the mothers). For such an effect, a large margin between exposure and RfD appears appropriate. However, given the large degree of uncertainty (a factor of around 160), a refined assessment might lead to a higher RfD, and this would be an option to be considered.

(b) Epididymis weight

The coverage of the deterministic RfD for epididymis weight (around 90%) might be considered sufficient (depending on the context of the assessment), and it is greater than, but quite close to, the deterministic estimate of human exposure. The probabilistic RfD (at 95% coverage) is lower than the estimated exposure by a factor of about 1.7. In this case, the degree of uncertainty is very large (approximately 580), and a refined assessment that reduces the uncertainty might result in a much higher RfD.

(c) Failure of sperm release

The coverage of the RfD for failure of sperm release is about 80%, and the deterministic RfD is again quite close to the deterministic estimate of the exposure of the human target population. The probabilistic RfD (with 95% coverage) in this case is even further below expected human exposure than for the other two end-points considered, indicating that the protection goal might not be achieved. Again, the degree of uncertainty in this case is very large, so there is a high potential that a higher-tier assessment would result in a higher RfD.

(d) Sensitivity analysis

Table A5.6 shows the APROBA output with respect to the relative contribution of the individual aspects to overall uncertainty.

**Table A5.6: Relative contributions of the individual aspects to overall uncertainty for the NOAEL-based assessments of further end-points**

<i>Aspect</i>	<i>Relative contribution (%) to overall uncertainty</i>	
	<i>Anomalies</i>	<i>Epididymis weight, sperm release</i>
NOAEL to BMD	36.9	23.7
Interspecies scaling	1.4	0.5
Interspecies TK/TD	18.6	11.9
Duration extrapolation	–	42.7
Intraspecies extrapolation	33.1	21.2
Data quality	10.0	–

BMD: benchmark dose; NOAEL: no-observed-adverse-effect level; TK/TD: toxicokinetics/toxicodynamics

For the two fertility end-points (epididymis weight, failure of sperm release), duration extrapolation is a dominating contributor. However, the uncertainty in this aspect could be reduced only by generating additional data.

In all three end-points, the NOAEL to BMD extrapolation is a major source of uncertainty. Therefore, the higher-tier assessment to be considered as the first option is one that reanalyses the dose–response data with a better statistical method: the BMD approach. This is clearly much cheaper and quicker than other options (e.g. producing additional experimental data).

(e) Additional uncertainties

The conclusion from the first-tier assessment was that a higher tier would be needed. Clearly, evaluating additional uncertainties (including secondary uncertainties) could only strengthen that conclusion. Therefore, at this stage, such an evaluation would not change the conclusion, and it will be omitted here.

**A5.4.2 Tier 2 (BMD-based) evaluation of uncertainties**

*A5.4.2.1 Prenatal development*

Appendix II gives a more detailed overview of the results from the BMD analyses of the three end-points selected from Khera et al. (1982), which are summarized only very briefly below.

(a) BMDL for resorptions

The BMD analysis of the combined results for resorptions in the two experiments reported by Khera et al. (1982) did not reveal significant differences among the dose–response relationships of both experiments (i.e. the first experiment using dose levels of 0, 5, 10 and 15 mg/kg bw per day and the second using 0, 0.5, 1 and 2.5 mg/kg bw per day). From the recommended set of quantal dose–response models, six models were accepted by the goodness-of-fit test, and the confidence intervals for the BMD were very similar among these models.

The range from the lowest BMDL to the highest BMDU for the BMD at an extra risk of 10% (BMR = 0.10) was from 2.6 to 3.5 mg/kg bw per day. The BMDL for resorptions was 2.6 mg/kg bw per day (in mice).

(b) BMDL for anomalous fetuses

In the case of fetal anomalies, most models indicated a significant difference between both experiments within Khera et al. (1982) regarding the potency parameter (*b*) of the models. The first study was found to result in lower BMDs than the second study. Seven models from the recommended set (with study as covariate) resulted in accepted fits.

The range from the lowest BMDL to the highest BMDU for the BMD at an extra risk of 10% was from 1.0 to 1.7 mg/kg bw per day. The BMDL for anomalous fetuses was 1.0 mg/kg bw per day (in mice).

(c) BMDL for fetal body weight

Both the exponential and the Hill models resulted in a significant difference in parameter *a* (reflecting background fetal weight). The exponential model indicated that in Collins et al. (2006), male and female rats differed in sensitivity to DON (parameter *b*), but the Hill model did not. Nonetheless, the confidence intervals for the BMD were similar between the two models.

The range from the lowest BMDL to the highest BMDU for the BMD at BMR = 0.05 was from 1.4 to 2.5 mg/kg bw per day. The BMDL for fetal weight was 1.4 mg/kg bw per day (in mice).

(d) Overall BMDL

The lowest BMDL was found for anomalous fetuses in mice: 1.0 mg/kg bw per day. The width of the confidence interval was a factor of 1.7.

*A5.4.2.2 Fertility*

Appendix II summarizes the results from the BMD analyses of the five reproduction endpoints selected from Sprando et al. (2005). Here, only the overall uncertainty ranges and BMDLs are given.

(a) BMDL for epididymal weight

The overall uncertainty range (based on analysis of left and right epididymis combined) was from 1.6 to 5.0 mg/kg bw per day. The BMDL was 1.6 mg/kg bw per day.

(b) BMDL for seminal vesicle weight

The overall uncertainty range was 1.1–3.2 mg/kg bw per day. The BMDL was 1.1 mg/kg bw per day.

(c) BMDL for testicular count

The overall uncertainty range was 1.8–6.7 mg/kg bw per day. The BMDL was 1.8 mg/kg bw per day.

(d) BMDL for germ cell degeneration

The overall uncertainty range was 0.28–2.2 mg/kg bw per day. The BMDL was 0.28 mg/kg bw per day.

(e) BMDL for failure of sperm release

The overall uncertainty range was 0.64–1.8 mg/kg bw per day. The BMDL was 0.64 mg/kg bw per day.

(f) Overall BMDL

The lowest BMDL was found for germ cell degeneration: 0.28 mg/kg bw per day. The width of the confidence interval was about a factor of 8.

*A5.4.2.3 Assessment factors and derivation of the reference dose*

The assessment factors to be applied for the various hazard characterization aspects involved are the same as those used in the tier 1 assessment based on a NOAEL (see Table A5.3).

Table A5.7 shows the derivation of RfDs by applying the assessment factors to the overall BMDLs for the identified effects. The RfDs derived in this way are somewhat higher than those based on the NOAEL approach for general toxicity and developmental effects, but not for male fertility. For the latter effect, the width of the BMD confidence interval was relatively large (factor of 8).

**Table A5.7: Deriving deterministic RfDs based on the BMDL approach**

	<i>Derivation of human RfDs (all values in µg/kg bw per day)</i>	
	<i>Development (anomalies in mice)</i>	<i>Reproduction (fertility of male rats)</i>
Dose–response analysis	BMDL = 1 000 CI width = 1.7	BMDL = 280 CI width = 8
Interspecies	100	28
Intraspecies	10	2.8
Data quality	<b>2<sup>a</sup></b>	–
Exposure duration	–	<b>0.28<sup>a</sup></b>

BMDL: lower confidence limit of the benchmark dose; CI: confidence interval; RfD: reference dose

<sup>a</sup> The final RfDs are printed in bold.

#### A5.4.2.4 Evaluation of uncertainties

The uncertainties for the above end-points of reproductive toxicity are again evaluated using the Microsoft Excel spreadsheet APROBA. The uncertainty distributions used for the aspects involved are the same as for the assessment based on the NOAEL in section A5.4.1, except that uncertainties associated with the NOAEL vanish and the uncertainties in the BMD derive from the specific data set considered rather than from a generic assumption.

The BMD approach selected fetal anomalies as the critical developmental end-point in line with the NOAEL approach. However, whereas the NOAEL approach for fertility resulted in the same NOAEL for five end-points, the BMD approach indicated that germ cell degeneration was the critical end-point in the fertility study. In the BMD-based assessment, germ cell degeneration will be considered to be a deterministic quantal effect – that is, the target human dose is defined as the dose at which 1% of the male subpopulation would experience the effect of mild germ cell degeneration.

The results of the uncertainty evaluation are summarized in Table A5.8 and are compared with the results for body weight obtained in section A5.3.

**Table A5.8: Summary of uncertainty evaluation of BMDL-based assessments**

	<i>Deterministic RfD (µg/kg bw per day)</i>	<i>Coverage (%)</i>	<i>Degree of uncertainty</i>	<i>Probabilistic RfD (µg/kg bw per day)</i>	<i>Estimated exposure in target population (µg/kg bw per day) (P95)</i>
Body weight reduction <sup>a</sup>	1.7	68	43	0.44	0.44
Development (fetal anomalies)	2	82	59	0.80	0.39
Fertility (mild germ cell degeneration)	0.28	91	375	0.16	0.84

bw: body weight; P95: 95th percentile; RfD: reference dose

<sup>a</sup> See Table A5.2.

Clearly, the degree of uncertainty in the target human dose has been considerably reduced (in particular for the anomalies), which was the primary aim of this higher-tier assessment based on the BMD. The probabilistic RfD for fetal anomalies increased compared with the tier 1

assessment, but only moderately. For the fertility end-point, the probabilistic RfD slightly decreased and therefore is still lower than human exposure, but the associated degree of uncertainty is still high as well.

*A5.4.2.5 Is a higher-tier assessment needed?*

Table A5.8 in the previous section summarized the results of the hazard characterization of DON, with the three end-points that were found to be most critical. The probabilistic RfD is equal to human exposure for general toxicity, somewhat higher than human exposure for developmental effects and about a factor of 4 lower than human exposure for fertility effects.

The latter is, however, combined with a high degree of uncertainty (factor of 375) – that is, the target human dose could well be much higher than the probabilistic RfD. Therefore, the risk manager might decide to demand a higher-tier assessment for fertility effects. Looking at the relative contributions of the individual uncertainties to overall uncertainty (results not shown), duration extrapolation is identified as the main driver of the uncertainty in the fertility  $HD_M^1$  (relative contribution: 49%). Thus, improving the database in this regard would be an obvious way to reduce uncertainty in the RfD for germ cell degeneration. A new study might at the same time decrease the PoD uncertainty (note the relatively large BMD confidence interval in this case).

The other question would be whether the 2-fold margin between the probabilistic RfD and human exposure for the developmental effect (anomalies) can be considered sufficient. Therefore, any additional uncertainties that have so far not been taken into account need to be evaluated. In addition, secondary uncertainties denoted as “case-specific” in Table 4.8 of the main text need to be addressed.

This is illustrated in Table A5.9, where the potential additional uncertainties are summarized, together with a quantitative evaluation for each aspect.

The quantitative evaluation shown in Table A5.9 is a sensitivity analysis, based on a number of recalculations with APROBA, aiming to answer the question: “What would the probabilistic RfD, and the degree of uncertainty, have been if alternative values or distributions had been used as input in the APROBA tool?” This will now be further discussed in more detail for each of the aspects in the table.

(a) Additional primary uncertainties

The first aspect in Table A5.9 addresses the statistical deficiency in deriving the BMD confidence interval, caused by not taking the clustering of observed anomalies within litters into account (that information was missing in Khera et al., 1982). Ignoring litter effects generally results in smaller confidence intervals, for instance, a factor of 1.5 or 2 narrower, based on previous experience with such data sets. A simple “sensitivity analysis” using the APROBA tool (by simply replacing the BMD confidence interval by a wider one) resulted in a small decrease of the probabilistic RfD (around 10–20%).

The second aspect relates to the body weight of the test animals used in allometric scaling. The mean body weight given in the study by Khera et al. (1982) relates to the dams’ weights upon mating. It might be argued that a higher body weight should have been used, to account for growth in the remaining study period and body weight gain during gestation. The APROBA calculation shows that doubling the body weight would result in an increase in the probabilistic RfD by around 25%.

**Table A5.9: Evaluation of additional uncertainties for the developmental effect (anomalies), based on sensitivity analysis using APROBA<sup>a</sup>**

<i>Aspect</i>	<i>Impact on probabilistic RfD</i>		
<b>Additional (primary) uncertainties</b>			
	<i>BMD confidence interval</i>	<i>Probabilistic RfD at 95% coverage (µg/kg bw per day)</i>	<i>Degree of uncertainty</i>
Litter effects in BMD CI not taken into account; could inflate the BMD CI by a factor of 1.5 or 2, based on incidental experience with other data sets	<b>(1000, 1700)</b> (500, 3400)	<b>0.80</b> 0.66	<b>59</b> 88
	<i>Assumed body weight of test animal (g)</i>	<i>Probabilistic RfD at 95% coverage (µg/kg bw per day)</i>	<i>Degree of uncertainty</i>
Default body weight of test animal used for allometric scaling	60 <b>30</b>	0.99 <b>0.80</b>	59 <b>59</b>
<b>(Case-specific) secondary uncertainties</b>			
	<i>Primary uncertainty (LCL, UCL) (µg/kg bw per day)</i>	<i>Probabilistic RfD at 95% coverage (µg/kg bw per day)</i>	<i>Degree of uncertainty</i>
TK/TD uncertainty distribution was based on general toxicity	(1/2, 2) <b>(1/3, 3)</b> (1/5, 5)	0.97 <b>0.80</b> 0.58	41 <b>59</b> 111
Database for intraspecies distribution did not include developmental effects	(4.4, 21) <b>(2.2, 42)</b> (1.1, 84)	1.23 <b>0.80</b> 0.47	26 <b>59</b> 178
Assumed distribution for data quality was based on expert judgement <sup>b</sup>	(1, 3) <b>(1, 5)</b> (1, 10)	1.13 <b>0.80</b> 0.41	50 <b>59</b> 81

BMD: benchmark dose; bw: body weight; CI: confidence interval; LCL: lower confidence limit; PoD: point of departure; RfD: reference dose; TK/TD: toxicokinetic/toxicodynamic; UCL: upper confidence limit

<sup>a</sup> The values used in the probabilistic analysis are printed in bold.

<sup>b</sup> For the data quality aspect, only the UCL has been varied. The LCL always has to be set at 1, because the same study without the deficiencies could have resulted in a lower, but not in a higher, PoD.

(b) Secondary uncertainties

For the aspects examined in the second part of Table A5.9, the LCL and UCL values (column 2) for the respective primary uncertainties have been varied to a moderate degree (i.e. by a factor of 2 up and down). Just as for the additional primary uncertainties, the impact on the probabilistic RfD and on the degree of uncertainty were again evaluated with APROBA.

In an individual hazard characterization, it is the responsibility of the hazard assessor to decide whether secondary uncertainty could indeed be so large as to assume a half or double confidence interval for the primary uncertainty of a particular aspect. For the DON case, the results given in Table A5.9 show that such changes in the primary uncertainties could correspond to a decrease in the probabilistic RfD of up to 50%, or to a somewhat smaller increase.

Although it would be unlikely that all of these additional uncertainties would apply simultaneously (to the extent evaluated), a 2-fold margin between RfD and exposure appears hardly sufficient. Further, it should be remembered that the target human dose in this case relates to a 10% extra risk of anomalies in 1% of the mothers, which may be considered an insufficient protection goal. Therefore, the risk manager might also consider demanding a higher-tier assessment for developmental effects. Although the relative contribution from intraspecies uncertainties is largest in this case (44%), it would be much easier to reduce the uncertainty related to poor data quality (contribution to uncertainty 25%) by demanding a new (good quality) developmental study for DON.

## A5.5 Conclusions and further options

Considering all critical end-points together, the results of the tier 2 hazard characterizations were as follows:

- For body weight reduction, there was no margin between exposure and the probabilistic RfD.
- For fetal anomalies, the margin between exposure and the probabilistic RfD was about 2-fold, which appeared inadequate considering additional uncertainties.
- For fertility end-points, the margin between exposure and the probabilistic RfD was less than 1.

Therefore, it was concluded that additional data would be needed to further refine the hazard characterization, aiming at a smaller uncertainty range of the  $HD_M^I$ . Given the relatively high costs of such data, it may be worthwhile to first explore other options. One option could be to perform a full probabilistic hazard characterization rather than the approximate probabilistic analysis by APROBA. Another option could be to evaluate the uncertainties underlying the exposure assessment. A better option would be to perform an IPRA in which the uncertainties in the exposure assessment and in the hazard characterization are taken into account at the same time. Such an approach makes the overall outcome less conservative than comparing the probabilistic results from both the hazard and exposure assessments performed separately. As another advantage, performing a full IPRA makes visible what aspects of the risk assessment as a whole are most uncertain. It might well be that reducing uncertainties in exposure would be much more effective than generating additional toxicity data. With a full picture of the relative contributions from all sources of uncertainty, a rational choice can be made of how to best use the available resources. For an illustration of a full IPRA applied to various substances (including DON), see Bokkers et al. (2009) or Slob et al. (2014).

## A5.6 References

Bokkers BGH, Bakker MI, Boon PE, Bos P, Bosgra S, Van der Heijden GWAM et al. (2009). The practicability of the integrated probabilistic risk assessment (IPRA) approach for substances in food. Bilthoven: National Institute for Public Health and the Environment (RIVM Report 320121001/2009; <http://www.rivm.nl/bibliotheek/rapporten/320121001.pdf>, accessed 18 November 2017).

Collins TF, Sprando RL, Black TN, Olejnik N, Eppley RM, Hines FA et al. (2006). Effects of deoxynivalenol (DON, vomitoxin) on in utero development in rats. *Food Chem Toxicol.* 44(6):747–57.

ECHA (2012). Guidance on information requirements and chemical safety assessment. Chapter R.8: Characterisation of dose [concentration]–response for human health. Version: 2.1. November 2012. Helsinki: European Chemicals Agency ([http://echa.europa.eu/documents/10162/13632/information\\_requirements\\_r8\\_en.pdf](http://echa.europa.eu/documents/10162/13632/information_requirements_r8_en.pdf), accessed 18 November 2017).

Iverson F, Armstrong C, Nera E, Truelove J, Fernie S, Scott P et al. (1995). Chronic feeding study of deoxynivalenol in B6C3F1 male and female mice. *Teratog Carcinog Mutagen.* 15(6):283–306.

Khera KS, Whalen C, Angers G, Vesonder RF, Kuiper-Goodman T (1982). Embryotoxicity of 4-deoxynivalenol (vomitoxin) in mice. *Bull Environ Contam Toxicol.* 29(4):487–91.

Pieters MN, Bakker M, Slob W (2004). Reduced intake of deoxynivalenol in The Netherlands: a risk assessment update. *Toxicol Lett.* 153:145–53.

Pieters MN, Freijer J, Baars AJ, Slob W (2001). Risk assessment of deoxynivalenol in food. An assessment of exposure and effects in the Netherlands. Bilthoven: National Institute for Public Health and the Environment (RIVM Report 388802 022; <http://www.rivm.nl/dsresource?objectid=83cccf06-d855-460a-8bb1-eb1d0f0c1431&type=org&disposition=inline>, accessed 18 November 2017).

Slob W, Bakker MI, Te Biesebeek JD, Bokkers BGH (2014). Exploring the uncertainties in cancer risk assessment using the integrated probabilistic risk assessment (IPRA) approach. *Risk Anal.* (in press).

Sprando RL, Collins TF, Black TN, Olejnik N, Rorie JI, Eppley RM et al. (2005). Characterization of the effect of deoxynivalenol on selected male reproductive endpoints. *Food Chem Toxicol.* 43(4):623–35.

Appendix I: Dose-response information considered in the case-study

**Table A5.10: Effects of DON on body weight of B6C3F1 mice following administration in the diet for 2 years**

<i>Group</i>	<i>Dose (mg/kg bw per day)</i>	<i>N</i>	<i>Body weight (g ± SD)<sup>a</sup></i>	<i>Relative to control (%)</i>
<b>Females</b>				
A	0	36	41.54 ± 6.26	0
B	0.12	42	38.71 ± 4.73**	-6.8
C	0.7	37	33.76 ± 3.92**	-8.7
D	1.5	38	28.55 ± 2.08**	-31.3
<b>Males</b>				
A	0	37	43.85 ± 2.69	0
B	0.1	35	43.51 ± 2.86	-0.8
C	0.5	43	40.04 ± 3.00**	-8.7
D	1.1	42	35.09 ± 2.56**	-20.0

bw: body weight; SD: standard deviation; \*\*:  $P < 0.01$

<sup>a</sup> Group means of the average body weight over lifetime.

Source: Iverson et al. (1995)

**Table A5.11: Frequency of resorptions in mice**

<i>Dose (mg/kg bw per day)</i>	<i>No. of pregnant females</i>	<i>Implants per pregnancy (mean ± SE)</i>	<i>No. of total implants<sup>a</sup></i>	<i>Resorptions/total implants (%)</i>	<i>No. of total resorptions<sup>b</sup></i>
<b>Study 1</b>					
0	15	16.9 ± 0.4	254	4	10
5	14	15.6 ± 0.5	218*	80*	174*
10	12	14.5 ± 1.0	174*	100*	174*
15	13	12.5 ± 1.3	162*	100*	162*
<b>Study 2</b>					
0	15	14.6 ± 0.6	219	5	11
0.5	16	15.7 ± 0.6	251	4	10
1	19	13.8 ± 0.8	262	3	8
2.5	17	15.0 ± 0.6	255*	9*	23*

bw: body weight; SE: standard error; \*:  $P < 0.05$

<sup>a</sup> Not given in the original study report, calculated from columns 2 and 3.

<sup>b</sup> Not given in the original study report, calculated from columns 4 and 5.

Source: Khera et al. (1982)

Table A5.12: Anomalous fetuses in mice

<i>Dose (mg/kg bw per day)</i>	<i>No. of fetuses examined</i>	<i>No. of anomalous fetuses (% examined)</i>
<b>Study 1</b>		
0	239	11 (4.6)
5	41	34 (82.9)*
<b>Study 2</b>		
0	207	8 (3.9)
0.5	236	9 (3.8)
1	254	22 (8.7)*
2.5	229	148 (64.6)*

\*:  $P < 0.05$  (Fisher exact test, one-sided)

Source: Khera et al. (1982)

Table A5.13: Fetal body weight in rats

<i>Dose (mg/kg bw per day)</i>	<i>No. of viable fetuses (mean <math>\pm</math> SEM)</i>	<i>Mean fetal body weight (g) (mean <math>\pm</math> SEM)</i>
<b>Males</b>		
0	7.4 $\pm$ 0.3	3.8 $\pm$ 0.03
0.5	6.6 $\pm$ 0.6	3.8 $\pm$ 0.03
1	6.3 $\pm$ 0.8	3.8 $\pm$ 0.03
2.5	7.5 $\pm$ 0.6	3.3 $\pm$ 0.03*
5	2.8 $\pm$ 0.7*	2.8 $\pm$ 0.05*
<b>Females</b>		
0	7.0 $\pm$ 0.4	3.6 $\pm$ 0.02
0.5	7.1 $\pm$ 0.6	3.6 $\pm$ 0.03
1	5.5 $\pm$ 0.6	3.5 $\pm$ 0.03
2.5	7.5 $\pm$ 0.4	3.2 $\pm$ 0.03*
5	2.9 $\pm$ 0.8*	2.6 $\pm$ 0.05*

bw: body weight; SEM: standard error of the mean; \*:  $P < 0.05$ 

Source: Collins et al. (2006)

Table A5.14: Male fertility end-points in rats<sup>a</sup>

<i>Dose (mg/kg bw per day)</i>	<i>No. of animals</i>	<i>Left epididymis weight</i>	<i>Right epididymis weight</i>	<i>Seminal vesicle weight</i>	<i>Testicular count/g testis</i>	<i>Germ cell degeneration</i>	<i>Failure of sperm release</i>
0	15	142.3 $\pm$ 3.8	150.5 $\pm$ 4.1	374.0 $\pm$ 8.8	141	1	0
0.5	15	146.7 $\pm$ 4.3	152.4 $\pm$ 5.0	378.9 $\pm$ 17.2	140	1	1
1	15	141.9 $\pm$ 4.0	154.2 $\pm$ 5.0	381.2 $\pm$ 14.6	137	3	0
2.5	15	136.8 $\pm$ 4.3*	139.9 $\pm$ 5.0*	331.4 $\pm$ 9.4*	137*	6*	11*
5	12	122.1 $\pm$ 5.7*	122.6 $\pm$ 5.7*	259.0 $\pm$ 15.2*	123*	8*	11*

bw: body weight; SEM: standard error of the mean; \*:  $P \leq 0.05$ <sup>a</sup> Organ weights are given as mg/g bw  $\pm$  SEM.

Source: Sprando et al. (2005)

Appendix II: Results of BMD analyses

Body weights (Iverson et al., 1995)

Table A5.15: Summary of results from fitting the exponential or Hill model to body weights as a function of dose<sup>a,b</sup>

Model	No. of model parameters	Log-likelihood	BMDL (mg/kg bw per day)	BMDU (mg/kg bw per day)
Exponential	7	307.52	0.17	0.33
Hill	7	307.43	0.18	0.34
<b>Overall uncertainty range</b>		<b>0.17–0.34</b>		

bw: body weight; BMDL: lower confidence limit of the benchmark dose; BMDU: upper confidence limit of the benchmark dose; BMR: benchmark response

<sup>a</sup> Data for both sexes were combined.

<sup>b</sup> BMR = 0.05.

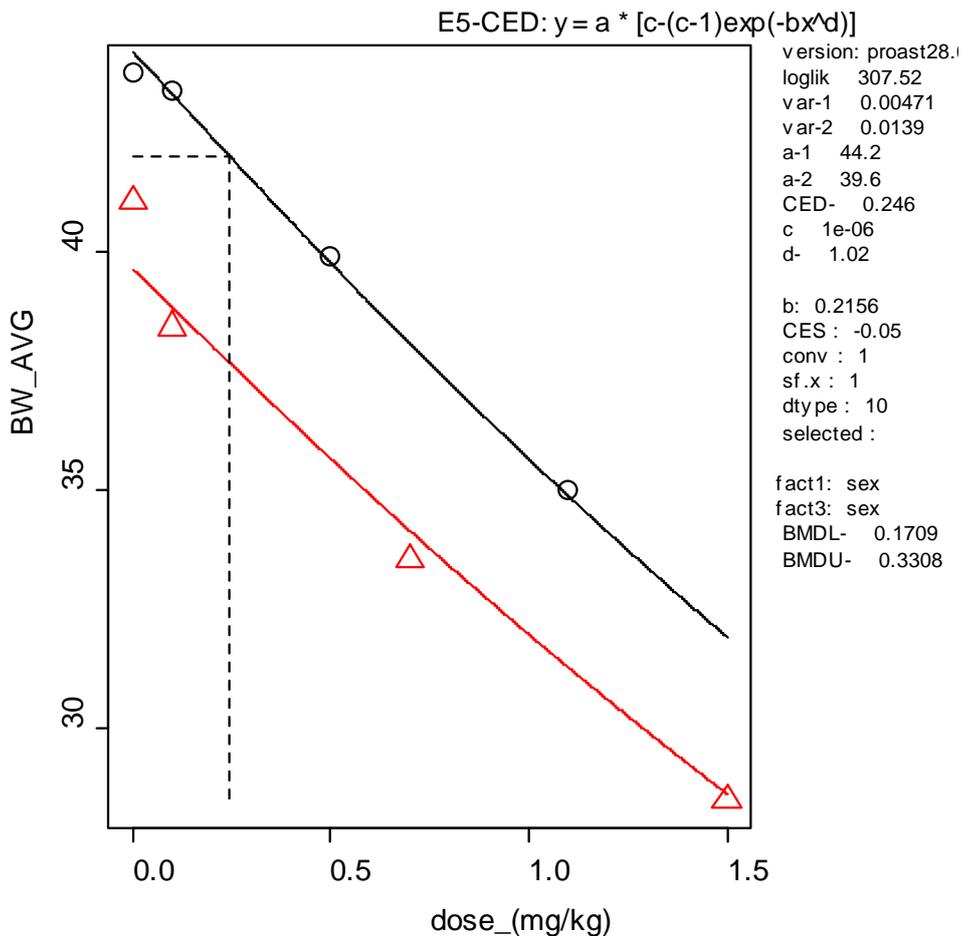


Fig. A5.10: Exponential model fitted to body weights as a function of dose. Circles and triangles denote the (geometric) means for males and females, respectively. The dashed line indicates the BMD<sub>05</sub> (approximately 240 µg/kg bw per day). In this model fit, parameter *b* (representing potency) was assumed equal for both sexes.

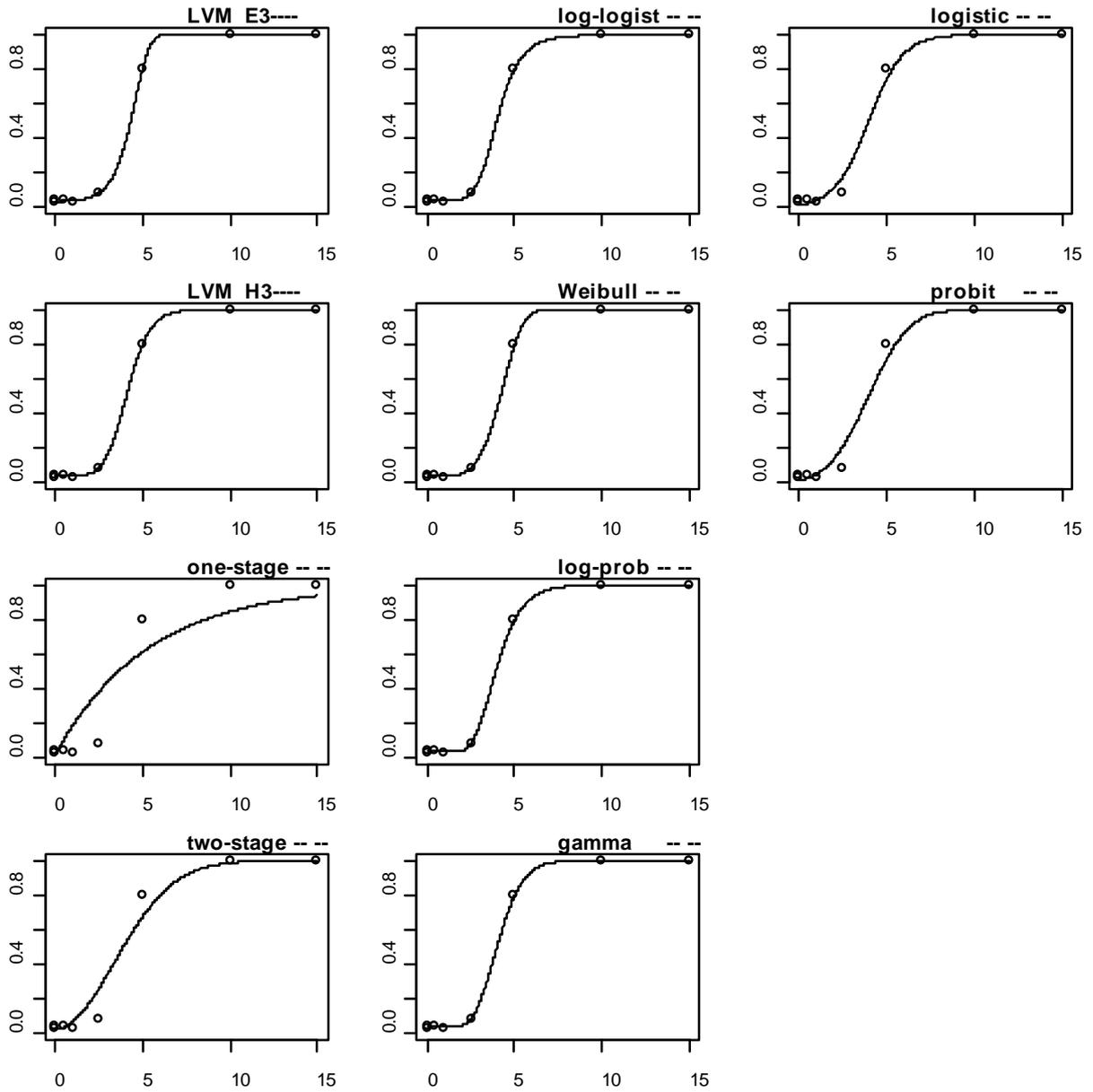
## Resorptions (Khera et al., 1982)

Table A5.16: Total resorptions – summary of dose–response analysis<sup>a</sup>

<i>Model</i>	<i>No. of model parameters</i>	<i>Log-likelihood</i>	<i>Accept</i>	<i>BMD</i>	<i>BMDL</i>	<i>BMDU</i>
Null	1	-1123.41	–	–	–	–
Full	8	-350.49	–	–	–	–
One-stage	2	-506.88	No	0.565	NA	NA
Two-stage	3	-390.62	No	1.55	NA	NA
Log-logistic	3	-351.75	Yes	2.84	2.61	3.15
Weibull	3	-351.13	Yes	2.87	2.59	3.22
Log-probit	3	-351.19	Yes	2.76	2.57	2.99
Gamma	3	-351.1	Yes	2.78	2.57	3.03
Logistic	2	-370.24	No	1.97	NA	NA
Probit	2	-379.27	No	1.73	NA	NA
LVM: E3-	3	-351.31	Yes	3	2.64	3.45
LVM: H3-	3	-351.14	Yes	2.85	2.59	3.18

BMD: benchmark dose; BMDL: lower confidence limit of the benchmark dose; BMDU: upper confidence limit of the benchmark dose; BMR: benchmark response; LVM: latent variable models, which are additionally available in the PROAST software (<http://www.proast.nl>); NA: not assessed

<sup>a</sup> BMR = 0.1; constraint: no; *P*-value goodness of fit: 0.05.



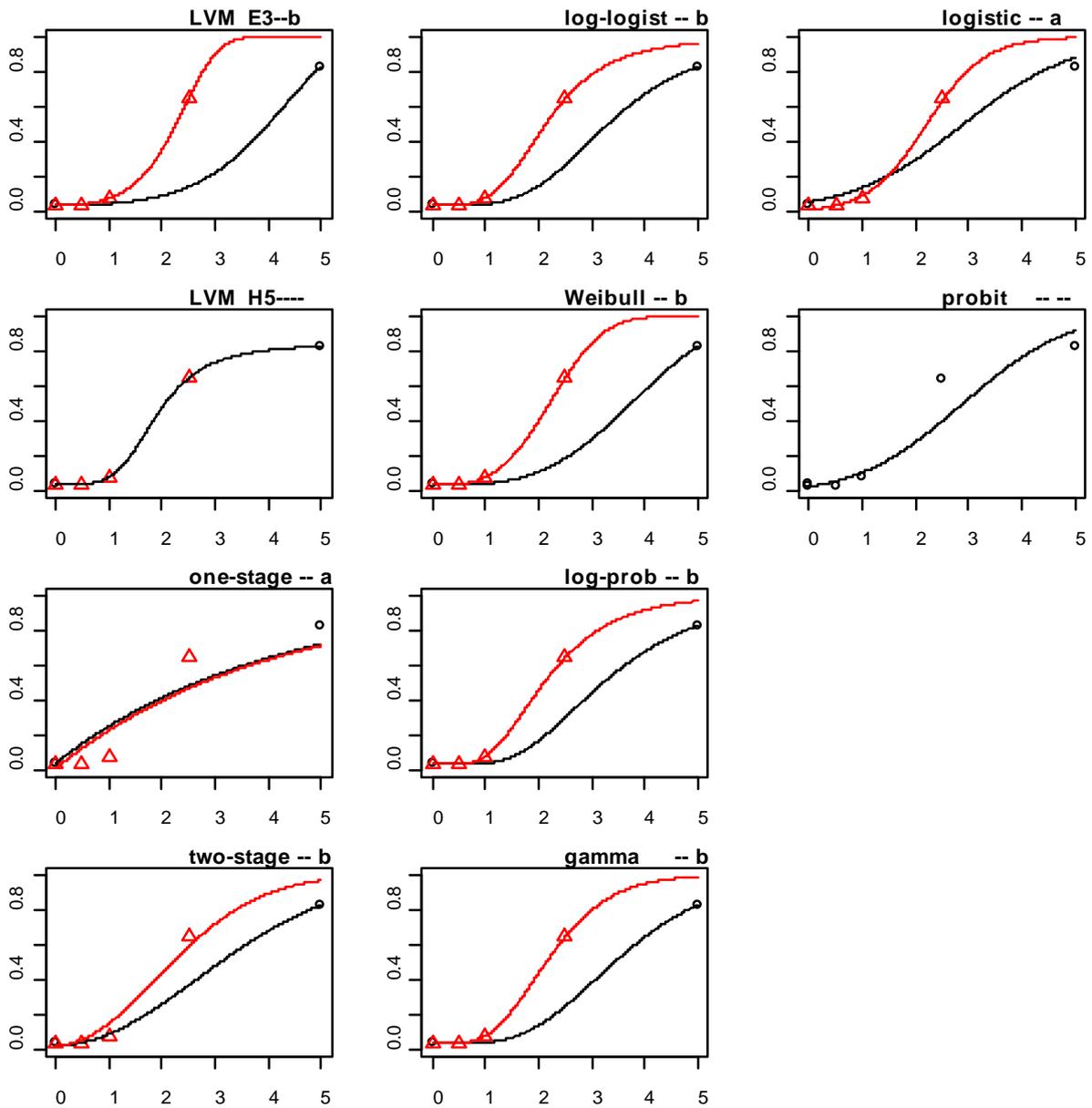
**Fig. A5.11:** Various models fitted to resorptions as a function of dose (Khera et al., 1982). The one-stage and two-stage models were rejected.

*Anomalous fetuses (Khera et al., 1982)***Table A5.17: Anomalous fetuses – summary of dose–response analysis<sup>a</sup>**

Model	Covar	No. of model parameters	Log-likelihood	Accept	BMD	BMDL	BMDU	Level
Null	–	1	–590.51	–	–	–	–	–
Full	–	6	–359.06	–	–	–	–	–
One-stage	A	3	–407.43	No	0.426	NA	NA	1
Two-stage	B	4	–368.38	No	0.874	NA	NA	2
Log-logistic	B	4	–359.26	Yes	1.23	1.07	1.46	2
Weibull	B	4	–359.30	Yes	1.27	1.09	1.52	2
Log-probit	B	4	–359.19	Yes	1.19	1.06	1.37	2
Gamma	B	4	–359.22	Yes	1.22	1.07	1.42	2
Logistic	A	3	–362.49	Yes	1.10	1.01	1.19	1
Probit	–	2	–391.45	No	1.13	NA	NA	1
LVM: E3-	B	4	–359.52	Yes	1.36	1.13	1.68	2
LVM: H5-	–	4	–359.25	Yes	1.21	1.06	1.43	1

BMD: benchmark dose; BMDL: lower confidence limit of the benchmark dose; BMDU: upper confidence limit of the benchmark dose; BMR: benchmark response; Covar: parameter that is found to be significantly different among studies; LVM: latent variable models, which are additionally available in the PROAST software (<http://www.proast.nl>); NA: not assessed

<sup>a</sup> BMR = 0.1; constraint: no; *P*-value goodness of fit: 0.05.



**Fig. A5.12: Various models fitted to anomalous fetuses (Khera et al., 1982).** Most models resulted in a significant difference between both studies (indicated by triangles or circles) within Khera et al. (1982).

Fetal body weights (Collins et al., 2006)

Table A5.18: Summary of results from fitting the exponential or Hill model to fetal body weights as a function of dose<sup>a,b</sup>

Model	Number of model parameters	Log-likelihood	BMDL	BMDU
Exponential	7	1109.46	1.59	2.45
Hill	6	1107.28	1.38	2.10
<b>Overall uncertainty range</b>		<b>1.38–2.45</b>		

BMDL: lower confidence limit of the benchmark dose; BMDU: upper confidence limit of the benchmark dose; BMR: benchmark response

<sup>a</sup> Data for both sexes were combined.

<sup>b</sup> BMR = 0.05.

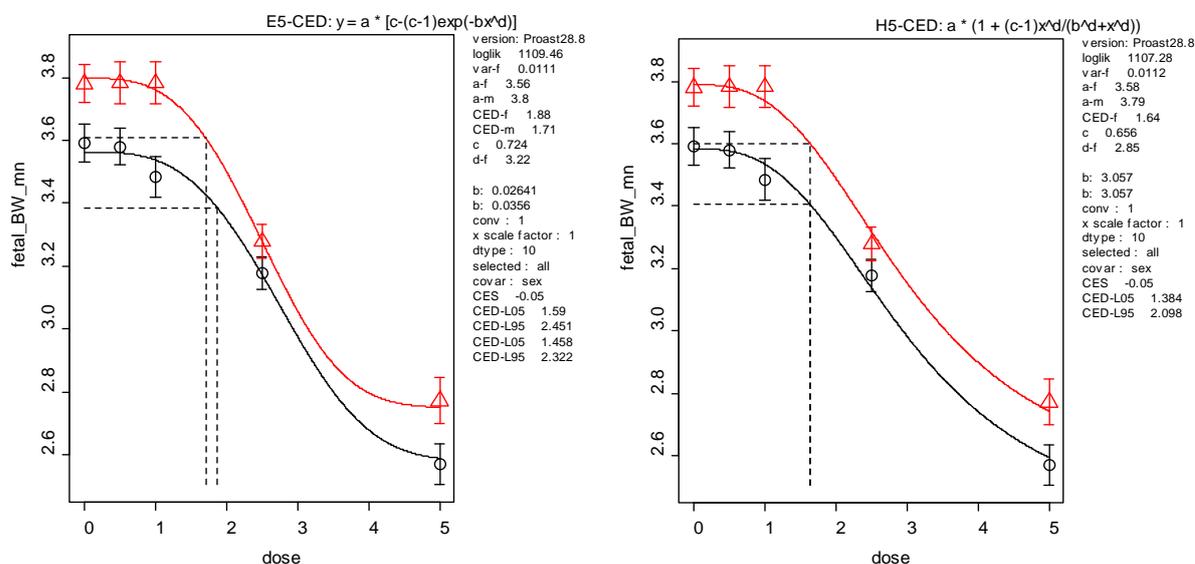


Fig. A5.13: Exponential model (left) and Hill model (right) fitted to fetal weights. Circles and triangles denote the (geometric) means for males and females, respectively. Dashed lines indicate the BMD<sub>05</sub>. The parameter *a* (background fetal weight) was found to differ significantly according to both models. Parameter *b* (representing potency) was found to differ among sexes in the exponential but not in the Hill model.

Epididymis weight (Sprando et al., 2005)

Table A5.19: Summary of results from fitting the exponential or Hill model to epididymis weight as a function of dose<sup>a</sup>

Model	Number of model parameters	Log-likelihood	BMDL	BMDU
<b>Left epididymis</b>				
Exponential	3	51.49	2.18	4.99
Hill	3	51.38	2.05	5.00
<b>Right epididymis</b>				
Exponential	3	45.73	1.73	3.32
Hill	3	45.53	1.58	3.30
<b>Overall uncertainty range</b>		<b>1.58–5.00</b>		

BMDL: lower confidence limit of the benchmark dose; BMDU: upper confidence limit of the benchmark dose; BMR: benchmark response

<sup>a</sup> BMR = 0.1.

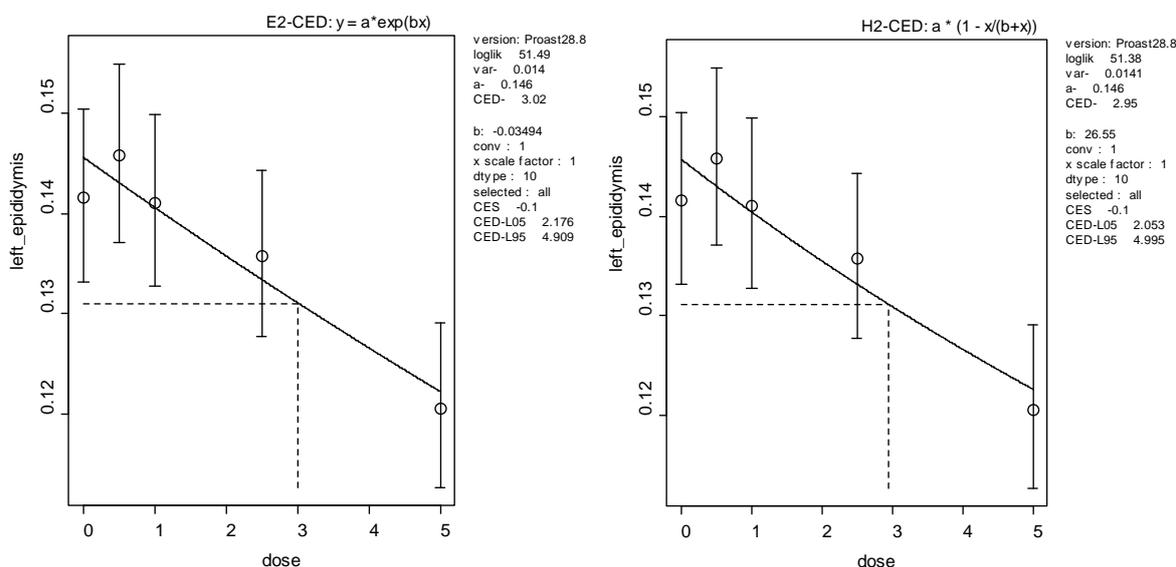


Fig. A5.14: Exponential model (left) and Hill model (right) fitted to left epididymis weight. Circles denote the (geometric) means. Dashed lines indicate the BMD<sub>10</sub>.

Seminal vesicle weight (Sprando et al., 2005)

Table A5.20: Summary of results from fitting the exponential or Hill model to seminal vesicle weights<sup>a</sup>

Model	Number of model parameters	Log-likelihood	BMDL	BMDU
Exponential	3	36.04	1.09	1.66
Hill	4	37.86	1.43	3.18
<b>Overall uncertainty range</b>		<b>1.09–3.18</b>		

BMDL: lower confidence limit of the benchmark dose; BMDU: upper confidence limit of the benchmark dose; BMR: benchmark response

<sup>a</sup> BMR = 0.1.

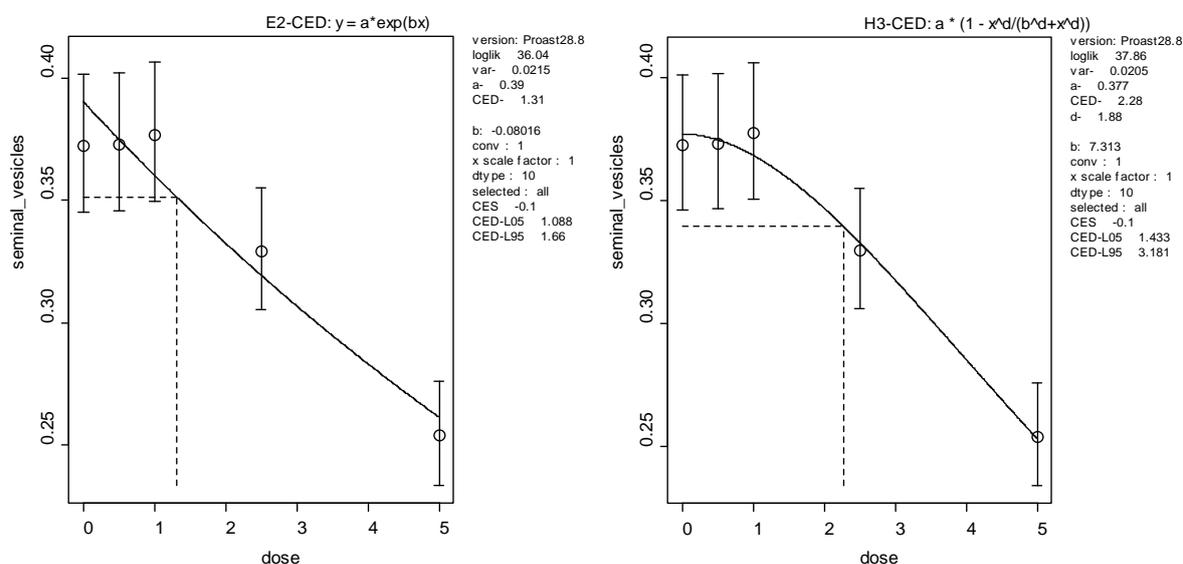


Fig. A5.15: Exponential model (left) and Hill model (right) fitted to seminal vesicle weight. Circles denote the (geometric) means. Dashed lines indicate the BMD<sub>10</sub>.

Testicular count (Sprando et al., 2005)

Table A5.21: Summary of results from fitting the exponential or Hill model to testicular counts<sup>a</sup>

Model	Number of model parameters	Log-likelihood	BMDL	BMDU
Exponential	3	27.18	1.94	6.42
Hill	3	27.11	1.81	6.67
<b>Overall uncertainty range</b>		<b>1.81–6.67</b>		

BMDL: lower confidence limit of the benchmark dose; BMDU: upper confidence limit of the benchmark dose; BMR: benchmark response

<sup>a</sup> BMR = 0.1.

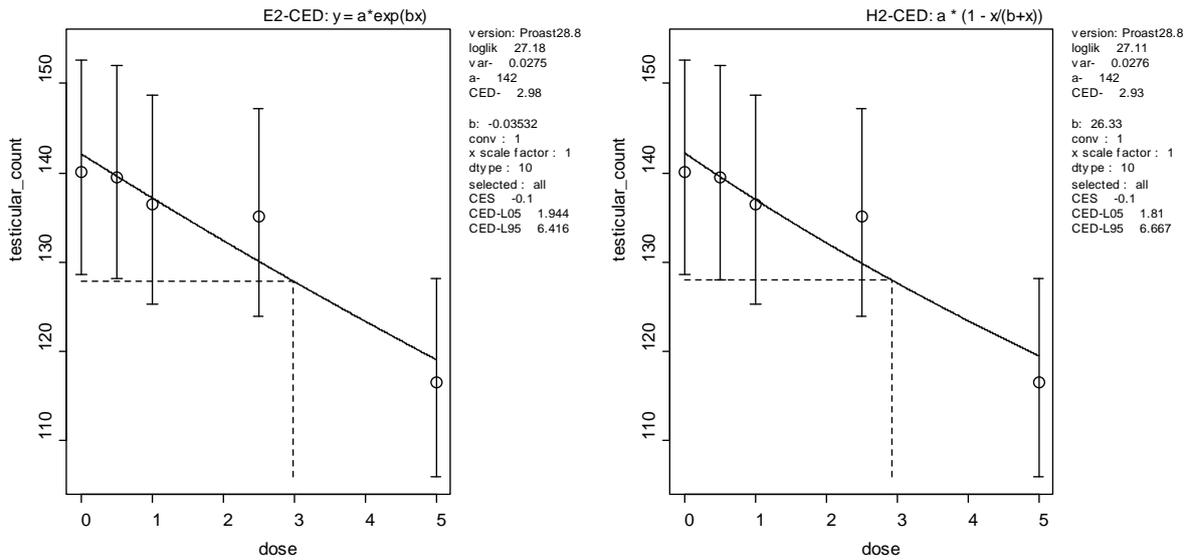


Fig. A5.16: Exponential model (left) and Hill model (right) fitted to testicular counts. Circles denote the (geometric) means. Dashed lines indicate the BMD<sub>10</sub>.

## Germ cell degeneration (Sprando et al., 2005)

Table A5.22: Summary of dose–response results for germ cell degeneration<sup>a</sup>

<i>Model</i>	<i>No. of parameters</i>	<i>Log-likelihood</i>	<i>Accept</i>	<i>BMD</i>	<i>BMDL</i>	<i>BMDU</i>
Null	1	-41.55	–	NA	NA	NA
Full	5	-32.59	–	NA	NA	NA
One-stage	2	-33.02	Yes	0.567	0.371	0.999
Two-stage	3	-32.84	Yes	0.790	0.382	1.93
Log-logistic	3	-32.74	Yes	0.895	0.278	2.14
Weibull	3	-32.78	Yes	0.852	0.297	2.13
Log-probit	3	-32.70	Yes	0.902	0.312	2.13
Gamma	3	-32.76	Yes	0.878	0.305	2.15
Logistic	2	-33.18	Yes	1.25	0.908	1.72
Probit	2	-33.09	Yes	1.16	1.01	1.43
LVM: E2-	2	-33.09	Yes	1.16	0.861	1.60
LVM: H2-	2	-32.91	Yes	0.648	0.429	1.06

BMD: benchmark dose; BMDL: lower confidence limit of the benchmark dose; BMDU: upper confidence limit of the benchmark dose; BMR: benchmark response; LVM: latent variable models, which are additionally available in the PROAST software (<http://www.proast.nl>); NA: not assessed

<sup>a</sup> BMR: 0.1; constraint: no; *P*-value goodness of fit: 0.05.

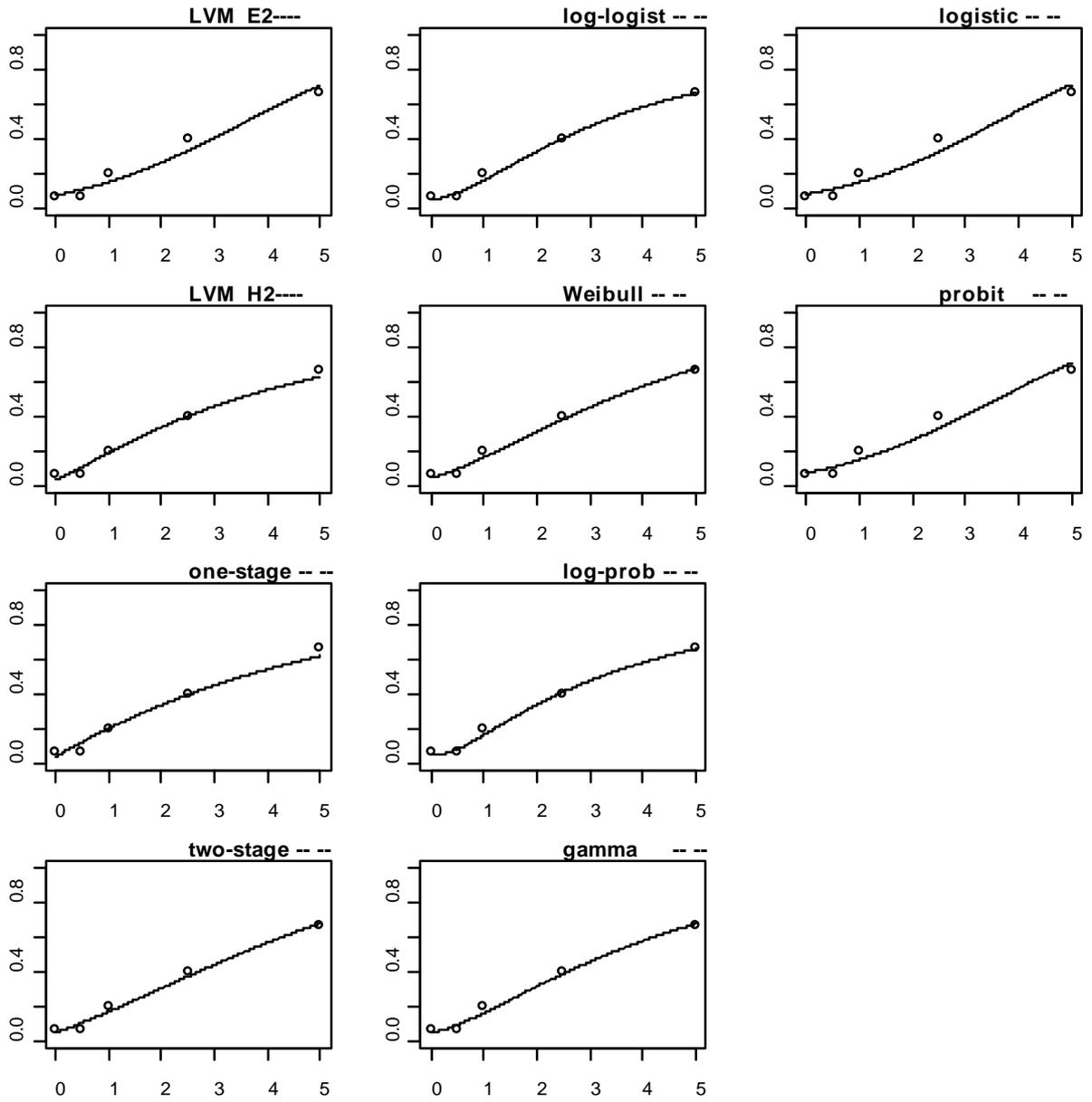


Fig. A5.17: Suite of models fitted to germ cell degeneration

## Failure of sperm release (Sprando et al., 2005)

Table A5.23: Summary of dose–response results for failure of sperm release

	<b>No. of parameters</b>	<b>Log- likelihood</b>	<b>Accept</b>	<b>BMD</b>	<b>BMDL</b>	<b>BMDU</b>
Null	1	-45.10	–	NA	NA	NA
Full	5	-15.81	–	NA	NA	NA
One-stage	2	-22.78	No	0.318	NA	NA
Two-stage	3	-19.22	No	0.879	NA	NA
Log-logistic	3	-18.10	Yes	1.27	0.685	1.75
Weibull	3	-19.21	No	0.848	NA	NA
Log-probit	3	-18.28	Yes	1.26	0.699	1.69
Gamma	3	-18.89	No	1.08	NA	NA
Logistic	2	-19.74	No	1.12	NA	NA
Probit	2	-20.26	No	1.06	NA	NA
LVM: E4-	3	-17.85	Yes	1.08	0.736	1.44
LVM: H3-	3	-18.55	Yes	1.07	0.635	1.57

BMD: benchmark dose; BMDL: lower confidence limit of the benchmark dose; BMDU: upper confidence limit of the benchmark dose; BMR: benchmark response; LVM: latent variable models, which are additionally available in the PROAST software (<http://www.proast.nl>); NA: not assessed

<sup>a</sup> BMR = 0.1; constraint: no; *P*-value goodness of fit: 0.05.

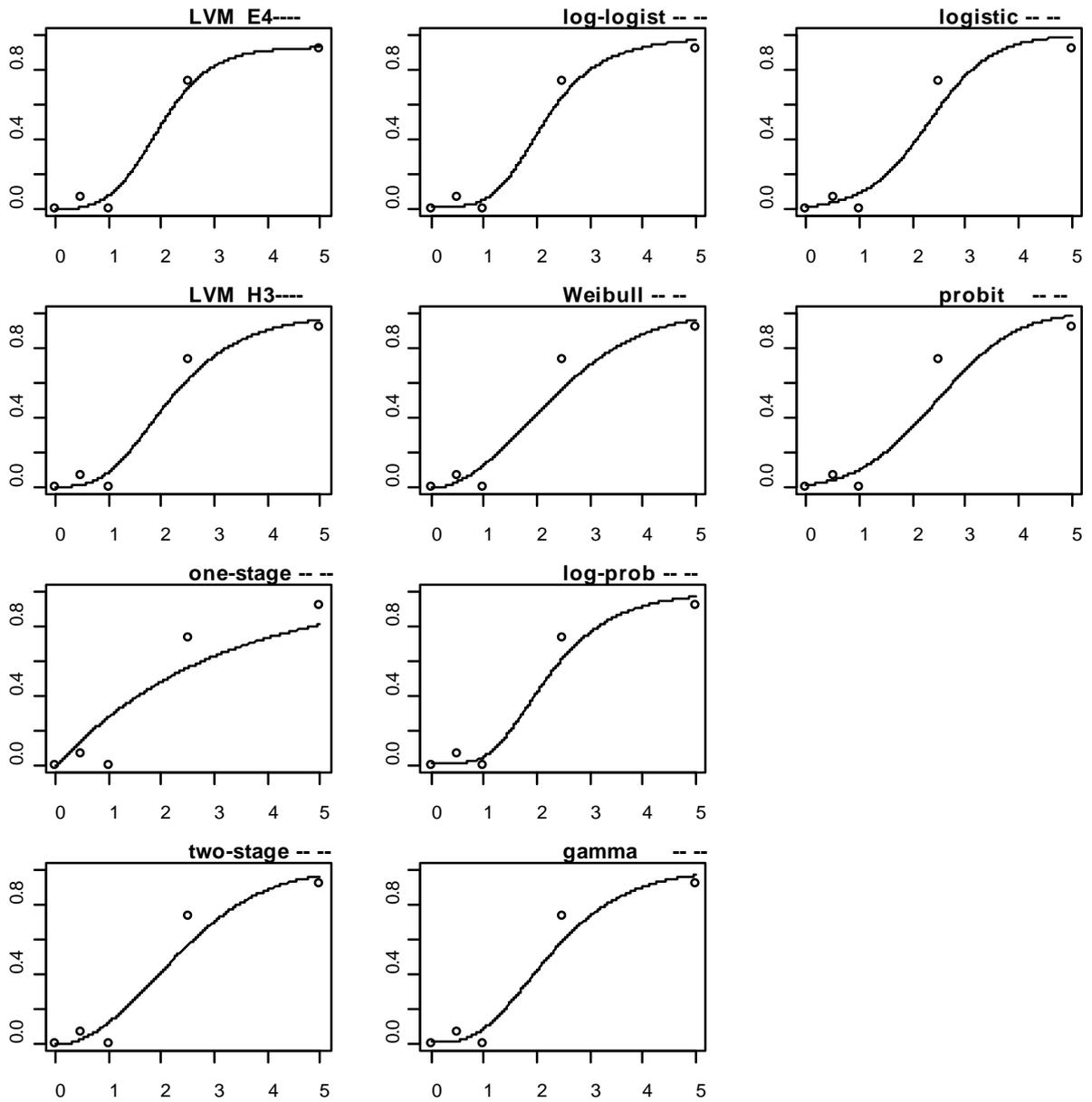


Fig. A5.18: Suite of models fitted to failure of sperm release

## THE HARMONIZATION PROJECT DOCUMENT SERIES

IPCS risk assessment terminology (No. 1, 2004)

Chemical-specific adjustment factors for interspecies differences and human variability: Guidance document for use of data in dose/concentration–response assessment (No. 2, 2005)

Principles of characterizing and applying human exposure models (No. 3, 2005)

Part 1: IPCS framework for analysing the relevance of a cancer mode of action for humans and case-studies; Part 2: IPCS framework for analysing the relevance of a non-cancer mode of action for humans (No. 4, 2007)

Skin sensitization in chemical risk assessment (No. 5, 2008)

Uncertainty and data quality in exposure assessment. Part 1: Guidance document on characterizing and communicating uncertainty in exposure assessment; Part 2: Hallmarks of data quality in chemical exposure assessment (No. 6, 2008)

Assessment of combined exposures to multiple chemicals: report of a WHO/IPCS international workshop on aggregate/cumulative risk assessment (No. 7, 2009)

WHO human health risk assessment toolkit: chemical hazards (No. 8, 2010)

Characterization and application of physiologically based pharmacokinetic models in risk assessment (No. 9, 2010)

Guidance for immunotoxicity risk assessment for chemicals (No. 10, 2012)

Guidance document on evaluating and expressing uncertainty in hazard characterization (No. 11, 2014)

*To order further copies of monographs in this series, please contact WHO Press,  
World Health Organization, 1211 Geneva 27, Switzerland  
(Fax No.: +41 22 791 4857; Email: [bookorders@who.int](mailto:bookorders@who.int)).  
The Harmonization Project Documents are also available on the web at  
<http://www.who.int/ipcs/en/>.*

ISBN 978-92-4-151354-8



9 789241 513548